Multi-View Stereo: A Tutorial

Yasutaka Furukawa
Washington University in St. Louis
furukawa@wustl.edu

Carlos Hernández
Google Inc.
carloshernandez@google.com

the essence of knowledge
Boston — Delft

Foundations and Trends® in Computer Graphics and Vision

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

Y. Furukawa and C. Hernández . *Multi-View Stereo: A Tutorial*. Foundations and Trends® in Computer Graphics and Vision, vol. 9, no. 1-2, pp. 1–148, 2013.

This Foundations and Trends® issue was typeset in LaTeX using a class file designed by Neal Parikh. Printed on acid-free paper.

ISBN: 978-1-60198-837-9

Foundations and Trends® in Computer Graphics and Vision
Volume 9, Issue 1-2, 2013

Editorial Board

Editorial Scope

Topics

Foundations and Trends® in Computer Graphics and Vision publishes survey and tutorial articles in the following topics:

- Rendering
- Shape
- Mesh simplification
- Animation
- Sensors and sensing
- Image restoration and enhancement
- Segmentation and grouping
- Feature detection and selection
- Color processing
- Texture analysis and synthesis
- Illumination and reflectance modeling
- Shape representation
- Tracking
- Calibration
- Structure from motion
- Motion estimation and registration
- Stereo matching and reconstruction
- 3D reconstruction and image-based modeling
- Learning and statistical methods
- Appearance-based matching
- Object and scene recognition
- Face detection and recognition
- Activity and gesture recognition
- Image and video retrieval
- Video analysis and event recognition
- Medical image analysis
- Robot localization and navigation

Information for Librarians

Foundations and Trends® in Computer Graphics and Vision, 2013, Volume 9, 4 issues. ISSN paper version 1572-2740. ISSN online version 1572-2759. Also available as a combined paper and online subscription.

Foundations and Trends® in
Computer Graphics and Vision
Vol. 9, No. 1-2 (2013) 1–148

DOI: 10.1561/0600000052

Multi-View Stereo: A Tutorial

Yasutaka Furukawa
Washington University in St. Louis
furukawa@wustl.edu

Carlos Hernández
Google Inc.
carloshernandez@google.com

Contents

Abstract

This tutorial presents a hands-on view of the field of multi-view stereo with a focus on practical algorithms. Multi-view stereo algorithms are able to construct highly detailed 3D models from images alone. They take a possibly very large set of images and construct a 3D plausible geometry that explains the images under some reasonable assumptions, the most important being scene rigidity. The tutorial frames the multi-view stereo problem as an image/geometry consistency optimization problem. It describes in detail its main two ingredients: robust implementations of photometric consistency measures, and efficient optimization algorithms. It then presents how these main ingredients are used by some of the most successful algorithms, applied into real applications, and deployed as products in the industry. Finally it describes more advanced approaches exploiting domain-specific knowledge such as structural priors, and gives an overview of the remaining challenges and future research directions.

Y. Furukawa and C. Hernández . *Multi-View Stereo: A Tutorial*. Foundations and Trends® in Computer Graphics and Vision, vol. 9, no. 1-2, pp. 1–148, 2013.
DOI: 10.1561/0600000052.

1

Introduction

Reconstructing 3D geometry from photographs is a classic Computer Vision problem that has occupied researchers for more than 30 years. Its applications range from 3D mapping and navigation to online shopping, 3D printing, computational photography, computer video games, or cultural heritage archival. Only recently however have these techniques matured enough to exit the laboratory controlled environment into the wild, and provide industrial scale robustness, accuracy and scalability.

Modeling the 3D geometry of real objects or scenes is a challenging task that has seen a variety of tools and approaches applied such as Computer Aided Design (CAD) tools [3], arm-mounted probes, active methods [110, 131, 11, 10] and passive image-based methods [162, 165, 176]. Among all, passive image-based methods, the subject of this tutorial, provide a fast way of capturing accurate 3D content at a fraction of the cost of other approaches. The steady increase of image resolution and quality has turned digital cameras into cheap and reliable high resolution sensors that can generate outstanding quality 3D content.

The goal of an image-based 3D reconstruction algorithm can be described as *"given a set of photographs of an object or a scene, estimate*

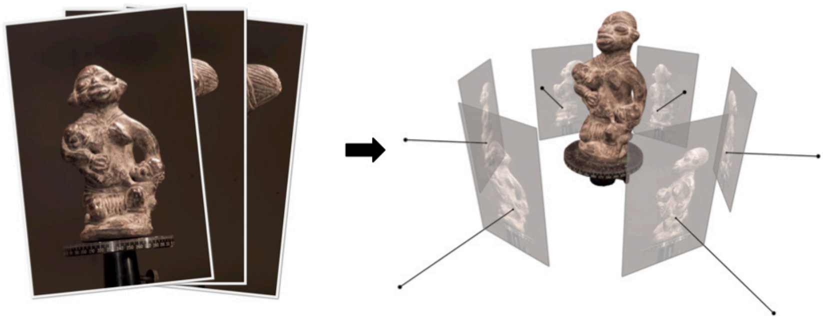

Figure 1.1: Image-based 3D reconstruction. Given a set of photographs (left), the goal of image-based 3D reconstruction algorithms is to estimate the most likely 3D shape that explains those photographs (right).

the most likely 3D shape that explains those photographs, under the assumptions of known materials, viewpoints, and lighting conditions" (See Figure 1.1). The definition highlights the difficulty of the task, namely the assumption that materials, viewpoints, and lighting are known. If these are not known, the problem is generally ill-posed since multiple combinations of geometry, materials, viewpoints, and lighting can produce exactly the same photographs. As a result, without further assumptions, no single algorithm can correctly reconstruct the 3D geometry from photographs alone. However, under a set of reasonable extra assumptions, e.g. rigid Lambertian textured surfaces, state-of-the-art techniques can produce highly detailed reconstructions even from millions of photographs.

There exist many cues that can be used to extract geometry from photographs: texture, defocus, shading, contours, and stereo correspondence. The latter three have been very successful, with stereo correspondence being the most successful in terms of robustness and the number of applications. Multi-view stereo (MVS) is the general term given to a group of techniques that use stereo correspondence as their main cue and use more than two images [165, 176].

All the MVS algorithms described in the following chapters assume the same input: a set of images and their corresponding camera parameters. This chapter gives an overview of an MVS pipeline starting from

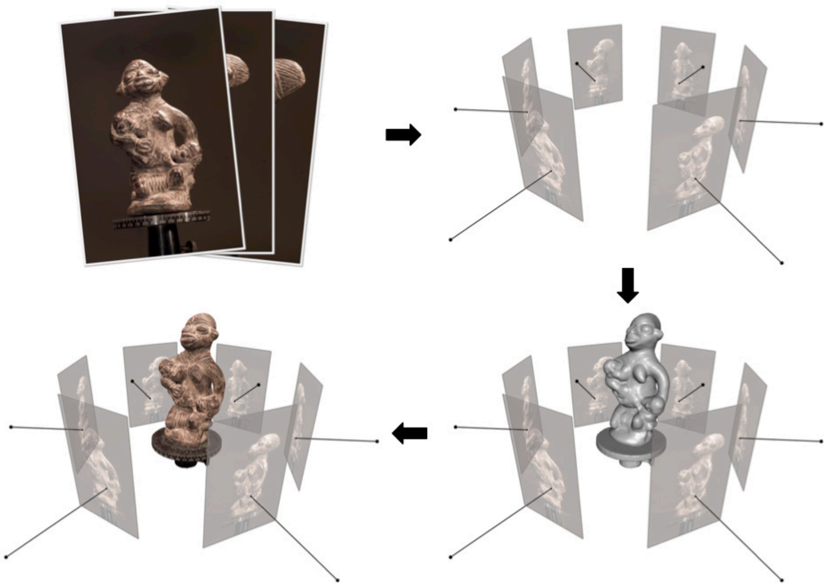

Figure 1.2: Example of a multi-view stereo pipeline. Clockwise: input imagery, posed imagery, reconstructed 3D geometry, textured 3D geometry.

photographs alone. An important take-home message of this chapter is simple: An MVS algorithm is only as good as the quality of the input images and camera parameters. Moreover, a large part of the recent success of MVS is due to the success of the underlying Structure from Motion (SfM) algorithms that compute the camera parameters.

Figure 1.2 provides a sketch of a generic MVS pipeline. Different applications may use different implementations of each of the main blocks, but the overall approach is always similar:

- Collect images,
- Compute camera parameters for each image,
- Reconstruct the 3D geometry of the scene from the set of images and corresponding camera parameters.
- Optionally reconstruct the materials of the scene.

Figure 1.3: Different MVS capture setups. From left to right: a controlled MVS capture using diffuse lights and a turn table, outdoor capture of small-scale scenes, and crowd-sourcing from online photo-sharing websites.

In the chapter we will give more insight into the first three main stages of MVS: imagery collection, camera parameters estimation, and 3D geometry reconstruction. Chapter 2 develops the notion of photo-consistency as the main signal being optimized by MVS algorithms. Chapter 3 presents and compares some of the most successful MVS algorithms. Chapter 4 discusses the use of domain knowledge, in particular, structural priors in improving the reconstruction quality. Chapter 5 gives an overview of successful applications, available software, and best practices. Finally Chapter 6 describes some of the current limitations of MVS as well as research directions to solve them.

1.1 Imagery collection

One can roughly classify MVS capture setups into three categories (See Figure 1.3):

- Laboratory setting,
- Outdoor small-scale scene capture,
- Large-scale scene capture using fleets or crowd-sourcing, e.g., cars, planes, drones, and Internet.

MVS algorithms first started in a laboratory setting [184, 147, 58], where the light conditions could be easily controlled and the camera could be easily calibrated, e.g. from a robotic arm [165], rotation table [93], fiducial markers [2, 43, 192], or early SfM algorithms [62]. MVS algorithms went through two major developments that took them to their current state: They left the laboratory setting to a small-scale outdoor scenes [174, 102, 85, 169, 190], e.g. a building facade or a fountain, then scaled up to much larger scenes, e.g. entire buildings and cities [129, 153, 97, 69].

These major changes were not solely due to the developments in the MVS field itself. It was a combination of new hardware to capture better images, more computation power, and scalable camera estimation algorithms.

Improvements in hardware: Two areas of hardware improvements had the most impact on MVS: digital cameras and computation power. Digital photography became mainstream and image digital sensors constantly improved in terms of resolution and quality. Additionally, mass production and miniaturization of geo positioning sensors (GPS) made them ubiquitous in digital cameras, tablets, and mobile phones. Although the precision of commercial units is not enough for MVS purposes, it does provide an initial estimate on camera parameters that can be refined using Computer Vision techniques. The second significant hardware improvement was computation power. The rise of inexpensive computer clusters [5] or GPU general computation [6] enabled SfM algorithms [25, 64] and MVS algorithms [69] to easily handle tens of thousands of images.

Improvements in Structure-from-Motion algorithms: Researchers have been working on visual reconstruction algorithms for decades [183, 182]. However, only relatively recently have these techniques matured enough to be used in large-scale industrial applications. Nowadays industrial algorithms are able to estimate camera parameters for millions of images. Two slightly different techniques have made great progress in recent years: Structure-from-Motion (SfM) [88] and Visual Simultaneous Localization and Mapping (VSLAM) [53]. Both rely on the correspondence cue and the assumption that the scene is rigid. SfM is most commonly used to compute camera models of unordered sets of images, usually offline, while VSLAM specializes in computing the location of a camera from a video stream, usually real-time. In this tutorial we focus on SfM algorithms, since a large majority of MVS algorithms are designed to work on unordered image sets, and rely on SfM to compute camera parameters. Note however that VSLAM has made very quick progress recently in the context of MVS [145, 180].

The term "camera parameters" refers to a set of values describing a camera configuration, that is, camera pose information consisting of location and orientation, and camera intrinsic properties such as focal length and pixel sensor size. There are many different ways or "models" to parameterize this camera configuration. In the following section, we discuss some of the most common camera projection models used in MVS applications.

1.2 Camera projection models

As mentioned in the introduction, MVS algorithms need additional knowledge in order to make the reconstruction problem well posed. In particular, MVS algorithms require that every input image has a corresponding camera model that fully describes how to project a 3D point in the world into a 2D pixel location in a particular image. The most commonly used camera model for MVS is the pinhole camera model, which is fully explained by a 3x4 projection camera matrix [88], defined up to a scale. This is the model commonly used with off-the-shelf digital cameras capturing still photographs. Any 3×4 projection matrix P can be decomposed into the product of a $3{\times}3$ upper triangular matrix K and a 3×4 pose matrix $[R|T]$

$$P = \underbrace{\begin{pmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix}}_{K} \cdot \left(\begin{array}{ccc|c} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{array}\right). \tag{1.1}$$

The matrix K is commonly referred to as the intrinsics matrix, because it is composed of quantities intrinsic to the camera: vertical and horizontal focal lengths (f_x, f_y), principal point (c_x, c_y), and skew s. The matrix $[R|T]$ is commonly known as the extrinsics matrix, where R is

Figure 1.4: Common deviations from pinhole camera model. Left: a fish eye lens exhibiting large radial distortion (top) and a rectified version of the same image after removing radial distortion (bottom). Right: rolling shutter artifacts caused by a fast moving object in the scene [155].

the rotation of the camera and T is the translation of the camera. Note that, due to the quality of digital sensors, one rarely estimates the 11 parameters of the projection matrix. In particular, pixels are assumed to have no skew ($s = 0$), and be square ($f_x = f_y$). Also, if an image has not been cropped, it is safe to assume the principal point is at the center of the image. As a result, a common pinhole camera model is just composed of 7 parameters: the focal length f, the rotation matrix R and the translation vector T.

If the attached lens is low quality, or wide-angle (See Figure 1.4 left), the pure pinhole model is not enough and often extended with a radial distortion model. Radial distortion is particularly important for high-resolution photographs, where small deviations from the pure pinhole model can amount to multiple pixels near the image boundaries.

Radial distortion can typically be removed from the photographs before they enter the MVS pipeline. If the radial distortion parameters of an image have been estimated, one can undistort the image by resampling as if it had been taken with an ideal lens without distortion (See

Figure 1.4 bottom left). Undistorting the images simplifies the MVS algorithm and often leads to faster processing times. Some cameras, e.g. those in mobile phones, incorporate dedicated hardware to remove radial distortion during the processing of the image just after its capture. Note however that rectifying wide-angle images will introduce resampling artifacts as well as field of view cropping. To avoid these issues MVS pipelines can support radial distortion and more complicated camera models directly, at the expense of extra complexity.

Finally, rolling shutter is another source of complexity particularly important for video processing applications (See Figure 1.4 right). A digital sensor with an electronic rolling shutter exposes each row of an image at slightly different times. This is in contrast to global shutters where the whole image is exposed at the same time. A rolling shutter often provides higher sensor throughput at the expense of a more complicated camera model. As a result, if the camera or the scene are moving while capturing the image, each row of the image captures effectively a slightly different scene. If the camera or scene motion is slow w.r.t. the shutter speed, rolling shutter effects can be small enough to be ignored. Otherwise the camera projection model needs to incorporate the effects [63].

1.3 Structure from Motion

There is a vast literature on Structure-from-Motion algorithms, and it is not our intention to throughly review it here. In the following, we will discuss some of the key aspects of SfM and how they relate to MVS algorithms.

SfM algorithms take as input a set of images and produce two things: the camera parameters of every image, and a set of 3D points visible in the images which are often encoded as tracks. A track is defined as the 3D coordinates of a reconstructed 3D point and the list of corresponding 2D coordinates in a subset of the input images. Most of the current state-of-the-art SfM algorithms share the same basic processing pipeline (See Figure 1.5):

- Detect 2D features in every input image.
- Match 2D features between images.
- Construct 2D tracks from the matches.
- Solve for the SfM model from the 2D tracks.
- Refine the SfM model using bundle adjustment.

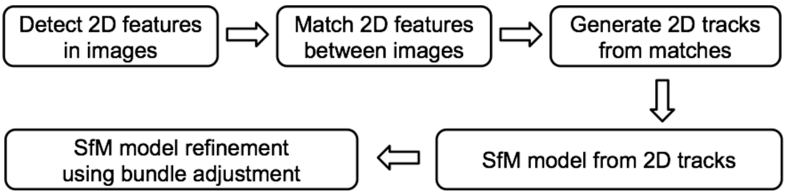

Figure 1.5: Main stages of a generic SfM pipeline, clockwise: feature detection, feature matching, track generation, structure-from-motion and bundle adjustment.

Initial work on SfM mainly focused on the geometry of two and three views under the assumption of a rigid scene [88]. Carlo Tomasi's technical perspective on visual reconstruction algorithms [182] presents an overview of the early work. One of the key developments for SfM was the use of RANSAC [61] to robustly estimate the epipolar geometry between two or three views given noisy matches.

Efforts then focused on two key components of the SfM algorithm: 1) computing a Euclidean reconstruction (up to a scale) from multiple cameras, that is, estimating both the camera parameters and 3D positions of the tracks, and 2) building longer 2D tracks. By the end of the 20th century, SfM algorithms were able to robustly compute models from large structured sets of images, e.g. from sequences of images or video sequences [62, 152] and the first SfM industrial solutions started to be commercialized for applications such as movie editing and special effects [4].

These initial systems were mainly designed for structured sets of images i.e., sets where the order of images matters, such as a video

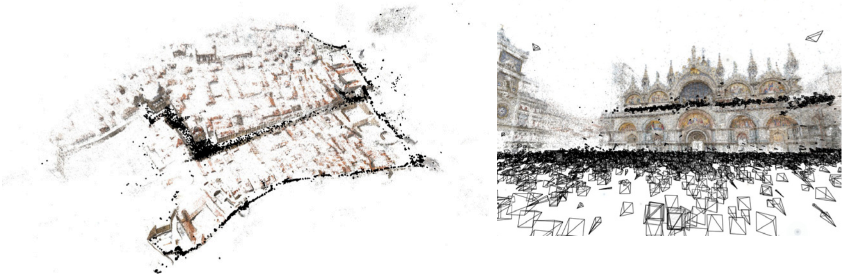

Figure 1.6: Large scale SfM examples from [25]. Left: SfM model of the city of Dubrovnik. Right: SfM model of San Marco Square in Venice.

sequence. Although some MVS applications can define such an order, for example, Google's StreetView [81] or Microsoft's Streetside [143], many recent MVS applications also use unordered sets of images captured at different times with different hardware, e.g. 3D maps from aerial images [108, 144, 30]. The development of fast and high quality feature detectors [87, 135, 57] and descriptors [135, 36, 159, 130, 26] was a crucial development towards making SfM work with unstructured datasets. High quality descriptors enabled building longer and higher quality tracks from images captured with very different pose and illumination.

The final ingredient to tackle large-scale SfM of unstructured photo collections was to improve the matching stage. In the case of unstructured photo collections, one does not have any prior knowledge of nearby candidate images that should be matched against. Therefore, every image has to be matched to every other image, which is computationally very expensive. Efficient indexing [146] combined with high quality descriptors allowed efficient pairwise matching of millions of images. Further work on simplifying the connectivity graph of the tracks [172] and parallelization [25, 64] lead to the current state-of-the-art SfM pipelines used in the industry, for example, Microsoft's photosynth [16] and Google's photo tours [15] (See Figure 1.6).

1.4 Bundle Adjustment

Although bundle adjustment [183] is not strictly a part of SfM, it is a very common step used to refine the initial SfM model. Given a set of camera parameters $\{P_i\}$, and a set of tracks $\{M^j, \{m_i^j\}\}$, where M^j denotes the 3D coordinate of a track, and m_i^j denotes the 2D image coordinate of its image projection in the i_{th} camera, bundle adjustment minimizes the following non-linear least squares error

$$E(P,M) = \sum_j \sum_{i \in V(j)} |P_i(M^j) - m_i^j|^2. \tag{1.2}$$

$V(j)$ is the list of camera indices where point M^j is visible, and $P_i(M^j)$ represents the projected 2D image coordinate of 3D point M^j in camera i using the camera parameters P_i.

$E(P,M)$ is typically measured in squared pixels, but a more common metric to express the accuracy of the estimation is to use the Root Mean Square Error or RMSE, which is measured in pixels and is defined as:

$$RMSE(P,M) = \sqrt{\frac{E(P,M)}{N}}, \tag{1.3}$$

where N is the number of residual terms being summed up in (1.2). Typical RMSE values before bundle adjustment are in the order of several pixels, while values after bundle adjustment are often sub-pixel.

The bundle adjustment framework enables the combination of multiple sensors with the SfM objective in a principled optimization framework. One way to fuse GPS and IMU constraints with SfM constraints is to simply add additional terms to (1.2) that penalize deviations of P_i from the predicted camera models from the GPS and IMU signals.

MVS algorithms are very sensitive to the accuracy of the estimated camera models. The reason is that, for efficiency purposes, they use the epipolar geometry defined by the camera models to restrict the 2D matching problem into a 1D matching problem (See Section 1.5 for more details). If the reprojection error is large, a pixel might never be compared against its real match, significantly degrading the MVS performance. The robustness of MVS to camera reprojection error depends mainly on how tolerant the matching criterion (namely the photo-consistency measures presented in Chapter 2) is to misalignments. Usually, the larger the domain Ω of the photo-consistency measure (See equation 2.1), the more robust the measure is. Unfortunately, large domains also tend to produce over smoothed geometry, so there is a compromise between accuracy and robustness.

Since MVS is so sensitive to reprojection errors, bundle adjustment is often a requirement for MVS, with the goal of sub-pixel reprojection errors. Note that, because reprojection error is measured in pixels, one can downsample the input images and rescale the camera parameters until the reprojection error drops below a certain threshold. This approach will work as long as the downsampled images still contain enough texture and details for MVS to work [72].

1.5 Multi-View Stereo

The origins of multi-view stereo can be traced back to human stereopsis and the first attempts to solve the stereoscopic matching problem as a computation problem [139]. Until this day, two-view stereo algorithms have been a very active and fruitful research area [162]. The multi-view version of stereo originated as a natural improvement to the two-view case. Instead of capturing two photographs from two different viewpoints, multi-view stereo would capture more viewpoints in-between to increase robustness, e.g. to image noise or surface texture [184, 147]. What started as a way to improve two-view stereo has nowadays evolved into a different type of problem.

Although MVS shares the same principles with such classic stereo algorithms, MVS algorithms are designed to deal with images with more varying viewpoints, such as an image set surrounding an object, and also deal with a very large number of images, even in the order of millions. The difference in the nature of the MVS problem ends up producing significantly different algorithms than the classic stereo counterpart. As an example, industrial applications for 3D mapping [108, 144, 30], process millions of photographs over hundreds of kilometers at a time, effectively reconstructing large metropolitan areas, countries and eventually the entire world.

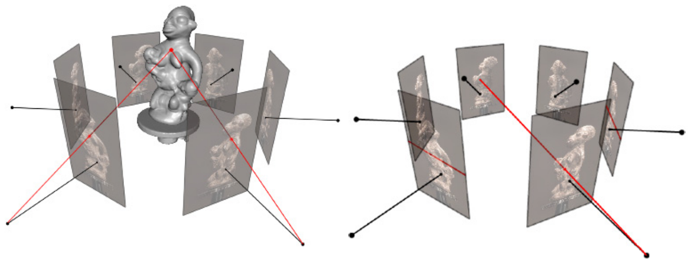

Figure 1.7: Matching images with known camera parameters. Left: The 3D geometry of the scene defines a correspondence between pixels in different images. Right: when camera parameters are known, matching a pixel in one image with pixels in another image is a 1D search problem.

Matching pixels across images is a challenging problem that is not unique to stereo or multi-view stereo. In fact, optical flow is another very active field in Computer Vision, tackling the problem of dense correspondence across images [33]. The main differences with MVS being that optical flow is typically a two image problem (similar to two-view stereo), cameras are not calibrated, and its main application is image interpolation rather than 3D reconstruction.

Note that in the case of MVS, where the camera parameters are known, solving for the 3D geometry of the scene is exactly equivalent to solving the correspondence problem across the input images. To see why, consider a 3D point belonging to the 3D scene geometry (See Figure 1.7 left). Projecting the 3D point into the set of visible cameras establishes a unique correspondence between the projected coordinates on each image.

Given a pixel in an image, finding the corresponding pixels in other images needs two ingredients:

- An efficient way to generate possible pixel candidates in other images.
- A measure to tell how likely a given candidate is the correct match.

If the camera geometry is not known, as is typically the case in optical flow, each pixel in an image can match any other pixel in another image. That is, for each pixel one has to do a 2D search in the other image. However, when the camera parameters are known (and the scene is rigid), the image matching problem is simplified from a 2D search to a 1D search (See Figure 1.7 right). A pixel in an image generates a 3D optic ray that passes through the pixel and the camera center of the image. The corresponding pixel on another image can only lie on the projection of that optic ray into the second image. The different geometric constraints that originate when multiple cameras look at the same 3D scene from different viewpoints are known as epipolar geometry [88].

As for measures to tell how likely a candidate match is, there is a vast literature on how to build so called *photo-consistency* measures that estimate the likelihood of two pixels (or groups of pixels) being in correspondence. Photo-consistency measures in the context of MVS are presented in more detail in Chapter 2.

References

[1] Acute3d. http://www.acute3d.com.

[2] Artoolkit. http://sourceforge.net/projects/artoolkit.

[3] Autodesk. http://en.wikipedia.org/wiki/Autodesk.

[4] Boujou. http://www.boujou.com.

[5] Cloud computing. http://en.wikipedia.org/wiki/Cloud_computing.

[6] Cuda: Compute unified device architecture. http://en.wikipedia.org/wiki/CUDA.

[7] Flickr. http://www.flickr.com.

[8] Google inc. http://www.google.com.

[9] Industrial light & magic. http://www.ilm.com.

[10] Kinect. http://en.wikipedia.org/wiki/Kinect.

[11] Lidar. http://en.wikipedia.org/wiki/Lidar.

[12] Multi-view environment. http://www.gris.informatik.tu-darmstadt.de/projects/multiview-environment.

[13] Openmvg (multiple view geometry). https://github.com/openMVG/openMVG.

[14] Panoramio. http://www.panoramio.com.

[15] Photo tours. http://google-latlong.blogspot.com/2012/04/visit-global-landmarks-with-photo-tours.html.

[16] Photosynth. `http://photosynth.net`.

[17] Picasa. `http://picasa.google.com`.

[18] Pix4d. `http://pix4d.com`.

[19] Primesense. `http://www.primesense.com`.

[20] Weta digital. `http://www.wetafx.co.nz`.

[21] Andrew Adams, Jongmin Baek, and Myers Abraham Davis. Fast high-dimensional filtering using the permutohedral lattice. In *Eurographics*, 2010.

[22] Y Adato, Y Vasilyev, O Ben-Shahar, and T Zickler. Toward a theory of shape from specular flow. In *IEEE International Conference on Computer Vision*, 2007.

[23] Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Ian Simon, Brian Curless, Steven M. Seitz, and Richard Szeliski. Building rome in a day. *Communications of the ACM*, 54(10):105–112, October 2011.

[24] Sameer Agarwal, Keir Mierle, and Others. Ceres solver. `https://code.google.com/p/ceres-solver/`.

[25] Sameer Agarwal, Noah Snavely, Ian Simon, Steven M. Seitz, and Richard Szeliski. Building Rome in a day. In *IEEE International Conference on Computer Vision*, 2009.

[26] A. Alahi, R. Ortiz, and P. Vandergheynst. Freak: Fast retina keypoint. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 510–517, June 2012.

[27] Neil Alldrin and David Kriegman. Toward reconstructing surfaces with arbitrary isotropic reflectance : A stratified photometric stereo approach. In *IEEE International Conference on Computer Vision*, 2007.

[28] Neil Alldrin and David Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[29] Pierre Alliez, David Cohen-Steiner, Yiying Tong, and Mathieu Desbrun. Voronoi-based variational reconstruction of unoriented point sets. In *Symposium on Geometry processing*, volume 7, pages 39–48, 2007.

[30] Apple. Apple maps. `http://www.apple.com/ios/maps`.

[31] Autodesk. 123d catch. `http://www.123dapp.com/catch`.

[32] C. Baillard, C. Schmid, A. Zisserman, and A. W. Fitzgibbon. Automatic line matching and 3D reconstruction of buildings from multiple views. In *ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery*, pages 69–80, 1999.

[33] Simon Baker, Daniel Scharstein, J. P. Lewis, Stefan Roth, Michael J. Black, and Richard Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92(1):1–31, March 2011.

[34] Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan Goldman. Patchmatch: a randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics*, 28(3):24, 2009.

[35] B.G. Baumgart. *Geometric modeling for computer vision.* PhD thesis, Stanford University, 1974.

[36] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346–359, June 2008.

[37] Thabo Beeler, Fabian Hahn, Derek Bradley, Bernd Bickel, Paul Beardsley, Craig Gotsman, Robert W. Sumner, and Markus Gross. High-quality passive facial performance capture using anchor frames. *ACM Transactions on Graphics*, 30:75:1–75:10, August 2011.

[38] S. Birchfield and C. Tomasi. Multiway cut for stereo and motion with slanted surfaces. In *IEEE International Conference on Computer Vision*, 1999.

[39] Stan Birchfield and Carlo Tomasi. A pixel dissimilarity measure that is insensitive to image sampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:401–406, 1998.

[40] Michael Bleyer, Christoph Rhemann, and Carsten Rother. Patchmatch stereo-stereo matching with slanted support windows. In *British Machine Vision Conference*, volume 11, pages 1–11, 2011.

[41] Matthew Bolitho, Michael Kazhdan, Randal Burns, and Hugues Hoppe. Multilevel streaming for out-of-core surface reconstruction. In *Symp. Geom. Proc.*, pages 69–78, 2007.

[42] Mario Botsch and Olga Sorkine. On linear variational surface deformation methods. *IEEE Transactions on Visualization and Computer Graphics*, 14(1):213–230, 2008.

[43] Jean-Yves Bouguet. Camera calibration toolbox for matlab. `http://www.vision.caltech.edu/bouguetj/calib_doc/`.

[44] Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, September 2004.

[45] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, November 2001.

[46] Derek Bradley, Wolfgang Heidrich, Tiberiu Popa, and Alla Sheffer. High resolution passive facial performance capture. *ACM Transactions on Graphics*, 29(4):41, 2010.

[47] Neill D.F. Campbell, George Vogiatzis, Carlos Hernández, and Roberto Cipolla. Automatic 3D object segmentation in multiple views using volumetric graph-cuts. In *British Machine Vision Conference*, volume 1, pages 530–539, 2007.

[48] Neill D.F. Campbell, George Vogiatzis, Carlos Hernández, and Roberto Cipolla. Using multiple hypotheses to improve depth-maps for multi-view stereo. In *10th European Conference on Computer Vision*, volume 5302 of *LNCS*, pages 766–779, 2008.

[49] Neill D.F. Campbell, George Vogiatzis, Carlos Hernández, and Roberto Cipolla. Automatic 3D object segmentation in multiple views using volumetric graph-cuts. *Image and Vision Computing*, 28(1):14 – 25, January 2010.

[50] Neill D.F. Campbell, George Vogiatzis, Carlos Hernández, and Roberto Cipolla. Automatic object segmentation from calibrated images. In *Visual Media Production (CVMP), 2011 Conference for*, pages 126 – 137, Nov. 2011.

[51] Guillermo D. Canas, Yuriy Vasilyev, Yair Adato, Todd Zickler, Steven Gortler, and Ohad Ben-Shahar. A linear formulation of shape from specular flow. In *IEEE International Conference on Computer Vision*, 2009.

[52] Brian Curless and Marc Levoy. A volumetric method for building complex models from range images. In *ACM SIGGRAPH*, 1996.

[53] A.J. Davison, I.D. Reid, N.D. Molton, and O. Stasse. Monoslam: Real-time single camera slam. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):1052–1067, June 2007.

[54] Paul E. Debevec, Camillo J. Taylor, and Jitendra Malik. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, ACM SIGGRAPH, pages 11–20, New York, NY, USA, 1996. ACM.

[55] Amaël Delaunoy and Emmanuel Prados. Gradient flows for optimizing triangular mesh-based surfaces: Applications to 3d reconstruction problems dealing with visibility. *International Journal of Computer Vision*, 95(2):100–123, November 2011.

[56] Amaël Delaunoy, Emmanuel Prados, P. Gargallo, J.-P. Pons, and P. Sturm. Minimizing the multi-view stereo reprojection error for triangular surface meshes. In *British Machine Vision Conference*, 2008.

[57] Tom Drummond Edward Rosten. Machine learning for high-speed corner detection. In *European Conference on Computer Vision*, pages 430–443. IEEE, 2006.

[58] O.D. Faugeras and R. Keriven. Variational principles, surface evolution, pde's, level-set methods, and the stereo problem. *IEEE Transactions on Image Processing*, 7(3):336–344, 1998.

[59] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.

[60] V. Ferrari, T. Tuytelaars, and L. Van Gool. Simultaneous object recognition and segmentation by image exploration. In *European Conference on Computer Vision*, 2004.

[61] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. *Commun. Assoc. Comp. Mach.*, 24:381–395, 1981.

[62] A. Fitzgibbon and A. Zisserman. Automatic camera recovery for closed or open image sequences. In *European Conference on Computer Vision*, volume I, pages 311–326, Freiburg, Germany, 1998.

[63] Forssén, Ringaby, and Hedborg. Computer vision on rolling shutter cameras. http://www.cvl.isy.liu.se/education/tutorials/rolling-shutter-tutorial, 2012.

[64] Jan-Michael Frahm, Pierre Fite-Georgel, David Gallup, Tim Johnson, Rahul Raguram, Changchang Wu, Yi-Hung Jen, Enrique Dunn, Brian Clipp, Svetlana Lazebnik, and Marc Pollefeys. Building rome on a cloudless day. In *European Conference on Computer Vision*, pages 368–381, Berlin, Heidelberg, 2010. Springer-Verlag.

[65] Simon Fuhrmann and Michael Goesele. Fusion of depth maps with multiple scales. *ACM Transactions on Graphics*, 30(6):148:1–148:8, December 2011.

[66] Simon Fuhrmann and Michael Goesele. Floating scale surface reconstruction. *ACM Transactions on Graphics*, 33(4):46, 2014.

[67] Yasutaka Furukawa, Brian Curless, Steven M. Seitz, and Richard Szeliski. Manhattan-world stereo. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[68] Yasutaka Furukawa, Brian Curless, Steven M. Seitz, and Richard Szeliski. Reconstructing building interiors from images. In *IEEE International Conference on Computer Vision*, 2009.

[69] Yasutaka Furukawa, Brian Curless, Steven M. Seitz, and Richard Szeliski. Towards Internet-scale multiview stereo. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

[70] Yasutaka Furukawa and Jean Ponce. Carved visual hulls for image-based modeling. *International Journal of Computer Vision*, 81(1):53–67, 2008.

[71] Yasutaka Furukawa and Jean Ponce. Dense 3d motion capture from synchronized video streams. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[72] Yasutaka Furukawa and Jean Ponce. Accurate camera calibration from multi-view stereo and bundle adjustment. *International Journal of Computer Vision*, 84(3):257–268, September 2009.

[73] Yasutaka Furukawa and Jean Ponce. Dense 3d motion capture for human faces. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1674–1681. IEEE, 2009.

[74] Yasutaka Furukawa and Jean Ponce. Accurate, dense, and robust multi-view stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8):1362–1376, August 2010.

[75] David Gallup, J-M Frahm, Philippos Mordohai, and Marc Pollefeys. Variable baseline/resolution stereo. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[76] David Gallup, Jan-Michael Frahm, Philippos Mordohai, Qingxiong Yang, and Marc Pollefeys. Real-time plane-sweeping stereo with multiple sweeping directions. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[77] David Gallup, Jan-Michael Frahm, and Marc Pollefeys. Piecewise planar and non-planar stereo for urban scene reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

[78] P. Gargallo, E. Prados, and P. Sturm. Minimizing the reprojection error in surface reconstruction from images. In *IEEE International Conference on Computer Vision*, pages 1–8, 2007.

[79] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S.M. Seitz. Multi-view stereo for community photo collections. In *IEEE International Conference on Computer Vision*, pages 1–8, 2007.

[80] Michael Goesele, Brian Curless, and Steven M. Seitz. Multi-view stereo revisited. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2402–2409, 2006.

[81] Google. Google maps. `http://www.google.com/maps/views/u/0/streetview`.

[82] Brian Gough. *GNU Scientific Library Reference Manual - Third Edition*. Network Theory Ltd., 3rd edition, 2009.

[83] Markus Gross and Hanspeter Pfister. *Point-Based Graphics*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2007.

[84] Saurabh Gupta, Ross Girshick, Pablo Arbeláez, and Jitendra Malik. Learning rich features from rgb-d images for object detection and segmentation. In *European Conference on Computer Vision*, pages 345–360. Springer, 2014.

[85] Martin Habbecke and Leif Kobbelt. A surface-growing approach to multi-view stereo reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[86] Bharath Hariharan, Pablo Arbeláez, Ross Girshick, and Jitendra Malik. Simultaneous detection and segmentation. In *European Conference on Computer Vision*, pages 297–312. Springer, 2014.

[87] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147–151, 1988.

[88] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.

[89] Samuel W. Hasinoff and Kiriakos N. Kutulakos. Confocal stereo. *International Journal of Computer Vision*, 81(1):82–104, 2009.

[90] Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. In *European Conference on Computer Vision*, ECCV'10, pages 1–14, Berlin, Heidelberg, 2010.

[91] C. Hernández, G. Vogiatzis, G. J. Brostow, B. Stenger, and R. Cipolla. Non-rigid photometric stereo with colored lights. In *IEEE International Conference on Computer Vision*, 2007.

[92] Carlos Hernández. *Stereo and Silhouette Fusion for 3D Object Modeling from Uncalibrated Images Under Circular Motion*. PhD thesis, Telecom Paris, Paris, France, 2004.

[93] Carlos Hernández and Francis Schmitt. Silhouette and stereo fusion for 3d object modeling. *Computer Vision and Image Understanding*, 96(3):367–392, 2004.

[94] Carlos Hernández, George Vogiatzis, and Roberto Cipolla. Probabilistic visibility for multi-view stereo. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[95] Carlos Hernández, George Vogiatzis, and Roberto Cipolla. Multiview photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3):548–554, 2008.

[96] Aaron Hertzmann and Steven M. Seitz. Example-based photometric stereo: Shape reconstruction with general, varying brdfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1254–1264, August 2005.

[97] V.H. Hiep, R. Keriven, P. Labatut, and J.-P. Pons. Towards high-resolution large-scale multi-view stereo. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1430–1437, June 2009.

[98] Heiko Hirschmuller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE Conference on Computer Vision and Pattern Recognition, pages 807–814, Washington, DC, USA, 2005. IEEE Computer Society.

[99] Heiko Hirschmuller and Daniel Scharstein. Evaluation of stereo matching costs on images with radiometric differences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9):1582–1599, 2009.

[100] Heiko HirschmÃijller. Evaluation of cost functions for stereo matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[101] Alexander Hornung, Er Hornung, and Leif Kobbelt. Robust and efficient photo-consistency estimation for volumetric 3d reconstruction. In *European Conference on Computer Vision*, pages 179–190, 2006.

[102] Alexander Hornung and Leif Kobbelt. Hierarchical volumetric multi-view stereo reconstruction of manifold surfaces based on dual graph embedding. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 503–510. IEEE, 2006.

[103] Alexander Hornung and Leif Kobbelt. Robust reconstruction of watertight 3 d models from non-uniformly sampled point clouds without normal information. In *Symposium on geometry processing*, pages 41–50. Citeseer, 2006.

[104] Alexander Hornung and Leif Kobbelt. Interactive pixel-accurate free viewpoint rendering from images with silhouette aware sampling. In *Computer Graphics Forum*, volume 28, pages 2090–2103. Wiley Online Library, 2009.

[105] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, and M. Gelautz. Fast cost-volume filtering for visual correspondence and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2):504–511, 2013.

[106] Xiaoyan Hu and P. Mordohai. A quantitative evaluation of confidence measures for stereo vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2121–2133, Nov 2012.

[107] Satoshi Ikehata and Kiyoharu Aizawa. Photometric stereo using constrained bivariate regression for general isotropic surfaces. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

[108] Google Inc. Google maps. `http://maps.google.com`.

[109] M. Jancosek and T. Pajdla. Multi-view reconstruction preserving weakly-supported surfaces. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

[110] R.A. Jarvis. A perspective on range finding techniques for computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5(2):122–139, March 1983.

[111] Dinghuang Ji, Enrique Dunn, and Jan-Michael Frahm. 3d reconstruction of dynamic textures in crowd sourced data. In *European Conference on Computer Vision*, 2014.

[112] Hailin Jin, Stefano Soatto, and Anthony J Yezzi. Multi-view stereo beyond lambert. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages I–171. IEEE, 2003.

[113] Hailin Jin, Stefano Soatto, and Anthony J Yezzi. Multi-view stereo reconstruction of dense shape and complex appearance. *International Journal of Computer Vision*, 63(3):175–189, 2005.

[114] Jörg H. Kappes, Bjoern Andres, Fred A. Hamprecht, Christoph Schnörr, Sebastian Nowozin, Dhruv Batra, Sungwoong Kim, Bernhard X. Kausler, Jan Lellmann, Nikos Komodakis, and Carsten Rother. A comparative study of modern inference techniques for discrete energy minimization problem. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2013. (accepted).

[115] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In *Symp. Geom. Proc.*, 2006.

[116] M. Kazhdan and H. Hoppe. Screened poisson surface reconstruction. *ACM Transactions on Graphics*, 32(3):29:1–29:13, July 2013.

[117] Changil Kim, Henning Zimmer, Yael Pritch, Alexander Sorkine-Hornung, and Markus Gross. Scene reconstruction from high spatio-angular resolution light fields. In *ACM SIGGRAPH*, 2013.

[118] Junhwan Kim, V. Kolmogorov, and R. Zabih. Visual correspondence using energy minimization and mutual information. In *IEEE International Conference on Computer Vision*, volume 2, pages 1033–1040, 2003.

[119] Kalin Kolev and Daniel Cremers. Integration of multiview stereo and silhouettes via convex functionals on convex domains. In *European Conference on Computer Vision*, 2008.

[120] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions using graph cuts. In *IEEE International Conference on Computer Vision*, volume 2, pages 508–515 vol.2, 2001.

[121] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In *European Conference on Computer Vision*, volume III, pages 82–96, Copenhagen, Denmark, 2002.

[122] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):147–159, 2004.

[123] Vladimir Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10):1568–1583, 2006.

[124] Adarsh Kowdle, Sudipta N Sinha, and Richard Szeliski. Multiple view object cosegmentation using appearance and stereo cues. In *European Conference on Computer Vision*, pages 789–803. Springer, 2012.

[125] Sanjiv Kumar and Martial Hebert. Discriminative random fields: A discriminative framework for contextual interaction in classification. In *IEEE International Conference on Computer Vision*, pages 1150–1157, 2003.

[126] Akash Kushal and Jean Ponce. A novel approach to modeling 3d objects from stereo views and recognizing them in photographs. In *European Conference on Computer Vision*, 2006.

[127] Avanish Kushal, Ben Self, Yasutaka Furukawa, David Gallup, Carlos Hernández, Brian Curless, and Steven M. Seitz. Photo tours. In *3DImPVT*, 2012.

[128] K.N. Kutulakos and S. M. Seitz. A theory of shape by space carving. *International Journal of Computer Vision*, 38(3):199–218, 2000.

[129] Patrick Labatut, Jean-Philippe Pons, and Renaud Keriven. Efficient multi-view reconstruction of large-scale scenes using interest points, delaunay triangulation and graph cuts. In *IEEE International Conference on Computer Vision*, 2007.

[130] S. Leutenegger, M. Chli, and R.Y. Siegwart. Brisk: Binary robust invariant scalable keypoints. In *IEEE International Conference on Computer Vision*, pages 2548–2555, Nov 2011.

[131] Marc Levoy, Kari Pulli, Brian Curless, Szymon Rusinkiewicz, David Koller, Lucas Pereira, Matt Ginzton, Sean Anderson, James Davis, Jeremy Ginsberg, Jonathan Shade, and Duane Fulk. The digital michelangelo project: 3d scanning of large statues. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, pages 131–144, 2000.

[132] Maxime Lhuillier and Long Quan. A quasi-dense approach to surface reconstruction from uncalibrated ima ges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):418–433, 2005.

[133] Yangyan Li, Xiaokun Wu, Yiorgos Chrysathou, Andrei Sharf, Daniel Cohen-Or, and Niloy J Mitra. Globfit: Consistently fitting primitives by discovering global relations. In *ACM Transactions on Graphics*, volume 30, page 52. ACM, 2011.

[134] William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, ACM SIGGRAPH, pages 163–169, New York, NY, USA, 1987. ACM.

[135] David G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the International Conference on Computer Vision-Volume 2 - Volume 2*, IEEE International Conference on Computer Vision, pages 1150–, Washington, DC, USA, 1999. IEEE Computer Society.

[136] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[137] Ziyang Ma, Kaiming He, Yichen Wei, Jian Sun, and Enhua Wu. Constant time weighted median filtering for stereo matching and beyond. In *IEEE International Conference on Computer Vision*, 2013.

[138] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *Medical Imaging, IEEE Transactions on*, 16(2):187–198, 1997.

[139] D. Marr and T. Poggio. A computational theory of human stereo vision. *Proceedings of Royal Society of London*, 204(1156):301–328, 1979.

[140] Kevin Matzen and Noah Snavely. Scene chronology. In *European Conference on Computer Vision*, 2014.

[141] Xing Mei, Xun Sun, Mingcai Zhou, Shaohui Jiao, Haitao Wang, and Xiaopeng Zhang. On building an accurate stereo matching system on graphics hardware. In *ICCV Workshops*, pages 467–474, 2011.

[142] Paul Merrell, Amir Akbarzadeh, Liang Wang, Jan michael Frahm, and Ruigang Yang David NistÃĺr. Real-time visibility-based fusion of depth maps. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[143] Microsoft. Bing maps. `http://blogs.bing.com/maps/2011/05/31/a-new-streetside-view/`.

[144] Microsoft. Bing maps. `http://www.bing.com/maps`.

[145] Richard A. Newcombe, S.J. Lovegrove, and A.J. Davison. Dtam: Dense tracking and mapping in real-time. In *IEEE International Conference on Computer Vision*, pages 2320–2327, 2011.

[146] David Nistér and Henrik Stewénius. Scalable recognition with a vocabulary tree. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2161–2168, 2006.

[147] M. Okutomi and T. Kanade. A multiple-baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(4):353–363, April 1993.

[148] Carl Olsson, Johannes Ulen, and Yuri Boykov. In defence of 3d-label stereo. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2013.

[149] G. P. Otto and T. K. W. Chau. 'region-growing' algorithm for matching of terrain images. *Image Vision Computing*, 7(2):83–94, 1989.

[150] Geoffrey Oxholm and Ko Nishino. Multiview shape and reflectance from natural illumination. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2163–2170. IEEE, 2014.

[151] Emanuel Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.

[152] M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. *International Journal of Computer Vision*, 32(1):7–25, 1999.

[153] M. Pollefeys, D. Nister, J.M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S.J. Kim, P. Merrell, et al. Detailed real-time urban 3d reconstruction from video. *International Journal of Computer Vision*, 78(2):143–167, 2008.

[154] Jean-Philippe Pons, Renaud Keriven, and Olivier Faugeras. Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score. *International Journal of Computer Vision*, 72(2):179–193, April 2007.

[155] Soren Ragsdale. Airplane prop + cmos rolling shutter, 2009.

[156] X. Ren and J. Malik. Learning a classification model for segmentation. In *IEEE International Conference on Computer Vision*, 2003.

[157] Christian Richardt, Douglas Orr, Ian Davies, Antonio Criminisi, and Neil A. Dodgson. Real-time spatiotemporal stereo matching using the dual-cross-bilateral grid. In Kostas Daniilidis, Petros Maragos, and Nikos Paragios, editors, *European Conference on Computer Vision*, volume 6313 of *Lecture Notes in Computer Science*, page 510âĂŞ523, September 2010.

[158] Sébastien Roy and Ingemar J. Cox. A maximum-flow formulation of the n-camera stereo correspondence problem. In *IEEE International Conference on Computer Vision*, 1998.

[159] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. Orb: An efficient alternative to sift or surf. In *IEEE International Conference on Computer Vision*, pages 2564–2571, Nov 2011.

[160] Szymon Rusinkiewicz and Marc Levoy. Qsplat: A multiresolution point rendering system for large meshes. In *ACM SIGGRAPH*, 2000.

[161] S. Savarese, M. Chen, and P. Perona. Local shape from mirror reflections. *International Journal of Computer Vision*, 64(1):31–67, 2005.

[162] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1/2/3):7–42, 2002.

[163] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47:7–42, 2002.

[164] Christopher Schroers, Henning Zimmer, Levi Valgaerts, AndrÃľs Bruhn, Oliver Demetz, and Joachim Weickert. Anisotropic range image integration. In *DAGM/OAGM Symposium'12*, pages 73–82, 2012.

[165] Steven M. Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 519–528, 2006.

[166] Steven M. Seitz and Charles R. Dyer. Photorealistic scene reconstruction by voxel coloring. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1067–, Washington, DC, USA, 1997. IEEE Computer Society.

[167] Ben Semerjian. A new variational framework for multiview surface reconstruction. In *European Conference on Computer Vision*, pages 719–734. Springer, 2014.

[168] Qi Shan, Changchang Wu, Brian Curless, Yasutaka Furukawa, Carlos Hernandez, and Steven M Seitz. Accurate geo-registration by ground-to-aerial image matching. In *3D Vision (3DV), 2014 2nd International Conference on*, volume 1, pages 525–532. IEEE, 2014.

[169] S.N. Sinha, P. Mordohai, and M. Pollefeys. Multi-view stereo via graph cuts on the dual of an adaptive tetrahedral mesh. In *IEEE International Conference on Computer Vision*, pages 1–8, 2007.

[170] Sudipta Sinha, Drew Steedly, and Richard Szeliski. Piecewise planar stereo for image-based rendering. In *IEEE International Conference on Computer Vision*, 2009.

[171] Sudipta N Sinha and Marc Pollefeys. Multi-view reconstruction using photo-consistency and exact silhouette constraints: A maximum-flow formulation. In *IEEE International Conference on Computer Vision*, 2005.

[172] N. Snavely, S. M. Seitz, and R. Szeliski. Skeletal graphs for efficient structure from motion. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[173] Noah Snavely. Bundler: Structure from motion (sfm) for unordered image collections. `http://www.cs.cornell.edu/~snavely/bundler`.

[174] C. Strecha, R. Fransens, and L. Van Gool. Wide-baseline stereo from multiple views: a probabilistic account. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2006.

[175] C. Strecha, T. Tuytelaars, and L. Van Gool. Dense matching of multiple wide-baseline views. In *IEEE International Conference on Computer Vision*, pages 1194–1201 vol.2, 2003.

[176] C. Strecha, W. Von Hansen, L. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2008.

[177] Christoph Strecha, Rik Fransens, and Luc Van Gool. Combined depth and outlier estimation in multi-view stereo. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2394–2401, Washington, DC, USA, 2006. IEEE Computer Society.

[178] Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer-Verlag New York, Inc., New York, NY, USA, 1st edition, 2010.

[179] Richard Szeliski, Ramin Zabih, Daniel Scharstein, Olga Veksler, Vladimir Kolmogorov, Aseem Agarwala, Marshall Tappen, and Carsten Rother. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(6):1068–1080, 2008.

[180] P. Tanskanen, K. Kolev, L. Meier, F. Camposeco Paulsen, O. Saurer, and M. Pollefeys. Live metric 3d reconstruction on mobile phones. In *IEEE International Conference on Computer Vision*, 2013.

[181] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *IEEE International Conference on Computer Vision*, pages 839–846, 1998.

[182] Carlo Tomasi. Visual reconstruction: Technical perspective. *Communications of the ACM*, 54(10):104–104, October 2011.

[183] B. Triggs, P.F. McLauchlan, Hartley R.I, and A.W. Fitzgibbon. Bundle adjustment - a modern synthesis. In *Vision Algorithms*, pages 298–372, 1999.

[184] R.Y. Tsai. Multiframe image point matching and 3-d surface reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5(2):159–174, March 1983.

[185] Ali Osman Ulusoy, Octavian Biris, and Joseph L Mundy. Dynamic probabilistic volumetric models. In *IEEE International Conference on Computer Vision*, pages 505–512. IEEE, 2013.

[186] Ali Osman Ulusoy and Joseph L Mundy. Image-based 4-d reconstruction using 3-d change detection. In *European Conference on Computer Vision*, 2014.

[187] Julien PC Valentin, Sunando Sengupta, Jonathan Warrell, Ali Shahrokni, and Philip HS Torr. Mesh based semantic modelling for indoor and outdoor scenes. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2067–2074. IEEE, 2013.

[188] Yuriy Vasilyev, Todd Zickler, Steven Gortler, and Ohad Ben-Shahar. Shape from specular flow: Is one flow enough? In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2561–2568, 2011.

[189] P. Viola and W.M.I.I.I. Wells. Alignment by maximization of mutual information. In *IEEE International Conference on Computer Vision*, pages 16–23, 1995.

[190] G. Vogiatzis, C. Hernández, P. H S Torr, and R. Cipolla. Multi-view stereo via volumetric graph-cuts and occlusion robust photo-consistency. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12):2241–2246, 2007.

[191] George Vogiatzis, P.H.S. Torr, and Roberto Cipolla. Multi-view stereo via volumetric graph-cuts. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

[192] George Vogiazis and Carlos Hernández. Automatic camera pose estimation from dot pattern. `http://george-vogiatzis.org/calib/`.

[193] H-H. Vu, P. Labatut, R. Keriven, and J.-P Pons. High accuracy and visibility-consistent dense multi-view stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(5):889–901, May 2012.

[194] Martin J. Wainwright, Tommi S. Jaakkola, and Alan S. Willsky. Map estimation via agreement on trees: message-passing and linear programming. *IEEE Transactions on Information Theory*, 51:2005, 2005.

[195] Sven Wanner and Bastian Goldluecke. Spatial and angular variational super-resolution of 4d light fields. In *European Conference on Computer Vision*, pages 608–621, 2012.

[196] O.J. Woodford, P.H.S. Torr, I. D. Reid, and A. W. Fitzgibbon. Global stereo reconstruction under second order smoothness priors. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[197] Robert J. Woodham. Shape from shading. chapter Photometric Method for Determining Surface Orientation from Multiple Images, pages 513–531. MIT Press, Cambridge, MA, USA, 1989.

[198] Changchang Wu. Visualsfm : A visual structure from motion system. `http://ccwu.me/vsfm`.

[199] Chenglei Wu, Bennett Wilburn, Yasuyuki Matsushita, and Christian Theobalt. High-quality shape from multi-view stereo and shading under general illumination. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 969–976. IEEE, 2011.

[200] Chenyang Xu and Jerry L. Prince. Gradient vector flow: A new external force for snakes. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 66–71. IEEE Computer Society, 1997.

[201] Shuntaro Yamazaki, Srinivasa G. Narasimhan, Simon Baker, and Takeo Kanade. The theory and practice of coplanar shadowgram imaging for acquiring visual hulls of intricate objects. *International Journal of Computer Vision*, 81(3):259–280, 2009.

[202] Ruigang Yang and M. Pollefeys. Multi-resolution real-time stereo on commodity graphics hardware. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages I–211–I–217 vol.1, June 2003.

[203] Ruigang Yang, Marc Pollefeys, Hua Yang, and Greg Welch. A unified approach to real-time, multi-resolution, multi-baseline 2d view synthesis and 3d depth estimation using commodity graphics hardware. *International Journal of Image and Graphics*, 4(04):627–651, 2004.

[204] Jonathan S. Yedidia, William T. Freeman, and Yair Weiss. Generalized belief propagation. In *IN NIPS 13*, pages 689–695. MIT Press, 2000.

[205] Kuk-Jin Yoon and In-So Kweon. Adaptive support-weight approach for correspondence search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):650–656, 2006.

[206] Ramin Zabih and John Woodfill. Non-parametric local transforms for computing visual correspondence. In *European Conference on Computer Vision*, pages 151–158, Stockhold, Sweden, 1994.

[207] C. Zach, T. Pock, and H. Bischof. A globally optimal algorithm for robust tv-l 1 range image integration. In *IEEE International Conference on Computer Vision*, 2007.

[208] Andrei Zaharescu, Edmond Boyer, and Radu Horaud. Transformesh: A topology-adaptive mesh-based approach to surface evolution. In *Asian Conference on Computer Vision*, 2007.

[209] L. Zebedin, J. Bauer, K. Karner, and H. Bischof. Fusion of feature- and area-based information for urban buildings modeling from aerial imagery. In *IEEE International Conference on Computer Vision*, pages IV: 873–886, 2008.

[210] Gang Zeng, Sylvain Paris, Long Quan, and Maxime Lhuillier. Surface reconstruction by propagating 3d stereo data in multiple 2d images. In *European Conference on Computer Vision*, 2004.

[211] Enliang Zheng, Ke Wang, Enrique Dunn, and Jan-Michael Frahm. Joint object class sequencing and trajectory triangulation (jost). In *European Conference on Computer Vision*, 2014.