V

Video Mosaicing

Zhigang Zhu
Department of Computer Science, Grove School of Engineering, The City College of The City University of New York, New York, NY, USA

Synonyms

Panoramic image generation; Video alignment and stitching; Video mosaicing; Video registration

Definition

Image mosaicing is the process of generating a composite image (mosaic) from a video sequence or in general from a set of overlapping images of a scene or an object, usually resulting in a mosaic image with a larger field of view than any of the original images.

Overview

When collecting a video about a scene or object, especially with a mobile platform (such as a robot), each individual image in the video may be limited compared to the desired final product, including limitations in the field of view, dynamic range, or image resolution. Generating mosaics with larger fields of view (Iketani et al. 2006; Irani et al. 1995; Rousso et al. 1998; Peleg and Ben-Ezra 1999; Shum and Szeliski 1999; Zhu et al. 2004), higher dynamic ranges (Eden et al. 2006), and/or higher image resolutions (Marzotto et al. 2004) facilitates video viewing, video understanding, video transmission, and video archiving. When the major objective of video mosaicing is to generate a complete (e.g., 360 degrees) view of an object (or a scene) by aligning and blending a set of overlapping images, the resulting image is also called a video panorama (Peleg and Ben-Ezra 1999; Shum and Szeliski 1999, 2000).

Key Research Findings

Video mosaicing takes in a video sequence and generates one or more mosaiced images with either a larger field of view, a higher dynamic range, a higher image resolution, or a combination of them. This entry will mainly discuss the principles in generating large field of view mosaics (panoramas), but similar principles can also be (mostly) applied to mosaics for other objectives (high dynamic range imaging and super-resolution imaging). Here, *video* mosaicing implies that the images in the sequence are taken by a video camera, usually at 30 frames per second, but images taken by a digital camera such that there is a large amount of spatial overlap between two consecutive frames can also be viewed as a video sequence.

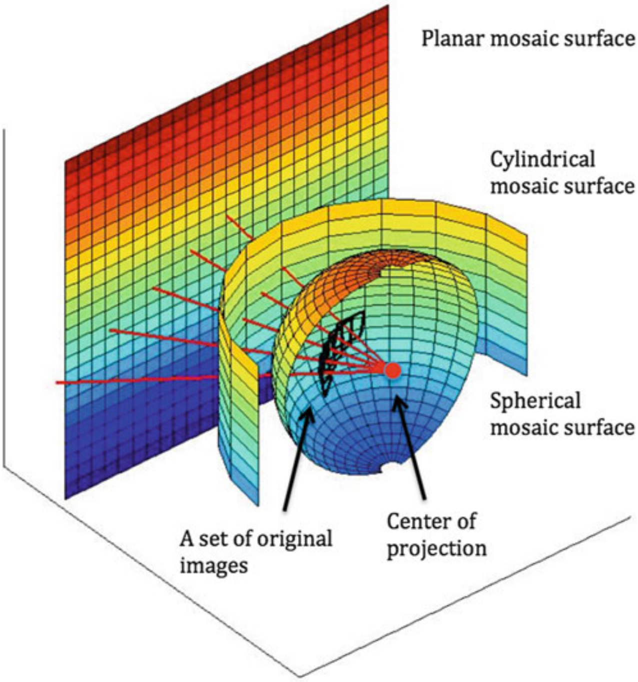
© Springer-Verlag GmbH Germany, part of Springer Nature 2020
M. H. Ang et al. (eds.), *Encyclopedia of Robotics*,
https://doi.org/10.1007/978-3-642-41610-1_106-1

There are three key components in a typical video mosaicing algorithm: motion modeling, image alignment, and image composition. Depending on the type of camera motion and the structure of the objects or scenes, the *motion model* can be a 2D rigid motion model (rotation, translation, scaling), an affine model, a perspective model (homography), or a full 3D motion model.

Popular video mosaicing methods (for a tutorial, please see Szeliski 2006) (e.g., Eden et al. 2006; Shum and Szeliski 2000) assume a pure rotation model of the camera in which the camera rotates around its center of projection (i.e., the optical center, sometimes called nodal point). In this case, the motion between two consecutive frames can be modeled by a homography, which is a 3 × 3 matrix. Then, depending on the fields of view (FOVs) of the mosaic, the projection model of the mosaic can be a planar perspective projection (FOV is less than 180°), a cylindrical projection (FOV is 360° in one direction), or a spherical projection (full 360° FOV in both direction). Figure 1 illustrates the relations between the original images and the three types of mapping surfaces each image can be projected onto: planar, cylindrical, and spherical.

However, the applications of video mosaics from a pure rotation camera are limited to mostly consumer applications such as personal photography, entertainment, and online maps. For more specialized robotic applications such as surveillance, remote sensing, navigation, and land planning, to name a few, the motion of the camera cannot be limited to a pure rotation. Translational motion usually cannot be avoided, causing the *motion parallax* problem to arise. Motion parallax is a monocular depth cue arising from the relative velocities of objects at various distances moving across the retina of a moving person. The term parallax refers to a change in position. Thus, in computer vision, motion parallax is a change in position of an object in images caused by the movement of the viewer (i.e., the camera). There are three kinds of treatments for the motion parallax problem. First, when the translational components are relatively small, the motion models can be approximated by a pure rotation. In this case, the generated mosaics lack geometric accuracy, but with some treatments for the small motion parallax and moving targets, such

Video Mosaicing, Fig. 1 Mapping a set of overlapping images into a mosaic: planar, cylindrical, or spherical



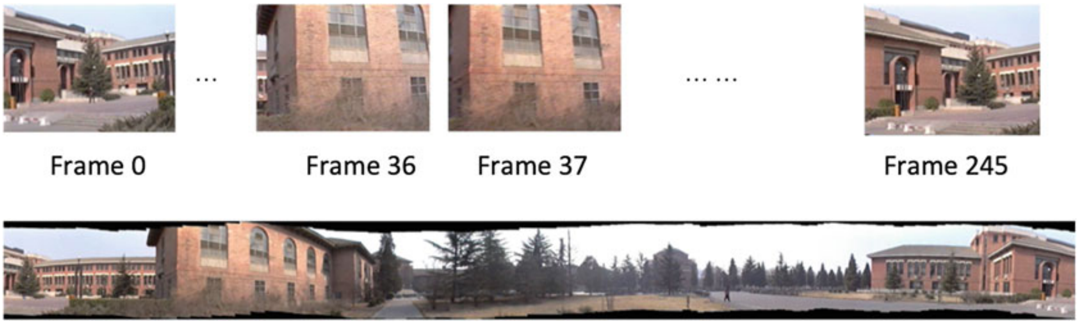
as the de-ghosting technique (Shum and Szeliski 2000), the mosaics generally look very good. Second, if the scene can be regarded as planar, for example, because the distance between the camera and the scene is much larger than the depth range of the scene, the perspective motion model (homography) or in some applications a 2D rigid motion model or an affine model can be used (Irani et al. 1995; Peleg et al. 2000; Zhu et al. 2005). In these cases, the problems are much simpler due to the 2D scene assumption. Finally, a 3D camera motion model is applied when the translational components of the camera motion are large and the scene is truly 3D. In this case motion parallax cannot be ignored or eliminated. Examples include a camera mounted on an airplane or a ground vehicle translating a large distance (Kumar et al. 1995; Rousso et al. 1998; Rademacher and Bishop 1998; Zhu et al. 2004) or a camera's optical center moving on a circular path (Peleg and Ben-Ezra 1999; Shum and Szeliski 1999). Here, multi-perspective projection models are used to generate the mosaics, enabling stereo mosaics or stereo panoramas to be created that preserve the 3D information in the scene, allowing the structure to be reconstructed and/or viewed in 3D. In this case, the accuracy of geometric modeling and image alignment is crucial for achieving the accuracy of 3D reconstruction and viewing.

Image alignment (or ***image registration***) is the process of finding the alignment parameters (e.g., the homography in the rotational case) between two consecutive images. Image alignment is a critical step in mosaic generation, for both seamless mosaicing and for accurate geometric representation. There are two approaches to image registration: direct methods or feature-based methods. In a direct method, usually a correlation approach is used to find the motion parameters. Here, the images are divided into small blocks, and each block in the first image is searched for over a predefined spatial range in the second image. The best match is determined by finding the maximal correlation value within the search range. Other approaches such as using optical flow or using an iterative optimization framework also belong to the direct methods, in which no explicit feature points are extracted. In a feature-based method, a feature detection operator such as the Harris corner (Harris and Stephens 1988) or SIFT (scale invariant feature transform) detector (Lowe 2004) is used first, and then the detected features are matched over the two frames to build up matches (Szeliski 2006). Either way, a parameter model is fitted using all the matches, usually using a robust parameter estimation method to eliminate erroneous feature matches. For more accurate or consistent results, a global optimization can be applied among more than two frames. For example, global alignment may be applied to all the frames in a full 360-degree circle in order to avoid gaps between the first and the last frame (Shum and Szeliski 2000).

Image composition is the step of combining aligned images together to form the viewable mosaic. There are three important issues in this step: compositing surface determination, coordinate transformation and image sampling, and pixel selection and blending. Mosaicing with the rotational camera model is a good starting point to discuss these issues (Fig. 1); mosaic compositing under other motion models is discussed afterward. An example of image composition from individual image frames is shown in Fig. 2 when the camera motion is rotational.

If the video sequence only has a few images, then one of the images can be selected as the reference image, and all the other images are warped and aligned with this reference image. In this case, the reference image with a perspective projection is the compositing surface, and therefore the final mosaic is a larger perspective image, which is an extension of the field of view of the reference image. However, this approach only works when the view angles of the images span less than 90°. If the camera rotates more than 90°, a cylindrical or a spherical surface should be selected as the compositing surface. A cylindrical surface is a good representation when a full 360 panoramic mosaic is to be generated, in one direction. And a spherical surface is suitable if 360 × 360 degree mosaics are to be created.

After a compositing surface is selected, the next issue is coordinate transformation and sampling. This is also called image warping. Given



Video Mosaicing, Fig. 2 A 360-degree panoramic mosaic generated on a cylindrical surface (the second row) from an image sequence of 246 frames captured by a rotational camera (four frames are shown on the first row) http://www-cs.engr.ccny.edu/~zhu/ThlibCylinder.JPG

the motion parameters obtained in the image registration step, the mapping between each frame to the final compositing surface can be calculated: for any pixel in an original image frame, its pixel location in the compositing surface can be calculated. For generating dense pixels, an interpolation schema is needed, such as nearest neighbor, bilinear, or cubic interpolation methods. Usually a backward mapping relation is utilized such that in the mapping area on the compositing surface, each pixel obtains a value from the original image frame, line by line and column by column. Therefore, for each integer pixel location in the mosaic, a decimal pixel location can be found in the original image; then an interpolation method is used in the original image to generate the value of the pixel in the mosaic.

The third important issue in image composition is pixel selection and blending. Naturally in generating mosaics, there are overlaps among consecutive frames, resulting in two key questions: First, *Where do we place the seam (*i.e.*, the stitching line)?* (the pixel selection problem). Second, *How do we select the values of pixels in the overlapping areas?* (the pixel blending problem). For the second problem, the simplest methods are to average all the pixels in the same location in the overlapping area or to use their median value. The former might create a so-called ghost effect due to moving objects, small motion parallax, or illumination changes, while the latter approach may generate a slightly better view effect. More sophisticated blending methods include Laplacian pyramid blending (Burt and Adelson 1983) and gradient domain blending (Agarwala et al. 2004). The pixel selection problem is important when moving objects or motion parallax exists in the scene. In these cases, to avoid a person being cut in half or appearing twice in the mosaic or to avoid cutting a 3D object that exhibits obvious motion parallax and hence could produce obvious misalignment in the mosaic, an optimal seam line can be selected at pixel locations where there are minimum misalignments between two frames (Eden et al. 2006).

Other considerations in image composition are high dynamic range imaging (Eden et al. 2006) and improved image resolution mosaicing (Marzotto et al. 2004). For the former, a composite mosaic represents larger dynamic ranges than individual frames using varying shutter speeds and exposures, while the latter uses the camera motion to generate higher spatial resolution in the mosaiced image than that of the original images.

Advanced Topics and Examples of Application

Video mosaicing finds applications with various robotic platforms such as for under-vehicle and pipe inspection (Dickson et al. 2002; Rzhanov 2013; Summan et al. 2015), aerial (e.g., UAV and drone) mapping (Taylor and Andersen 2008; Colorado et al. 2015; Xu et al. 2016), and ground robot localization [Zheng and Tsuji 1992; Zhu and Hanson 2004]. Examples of applications include personal video capture (Agarwala et al. 2004; Marzotto et al. 2004;

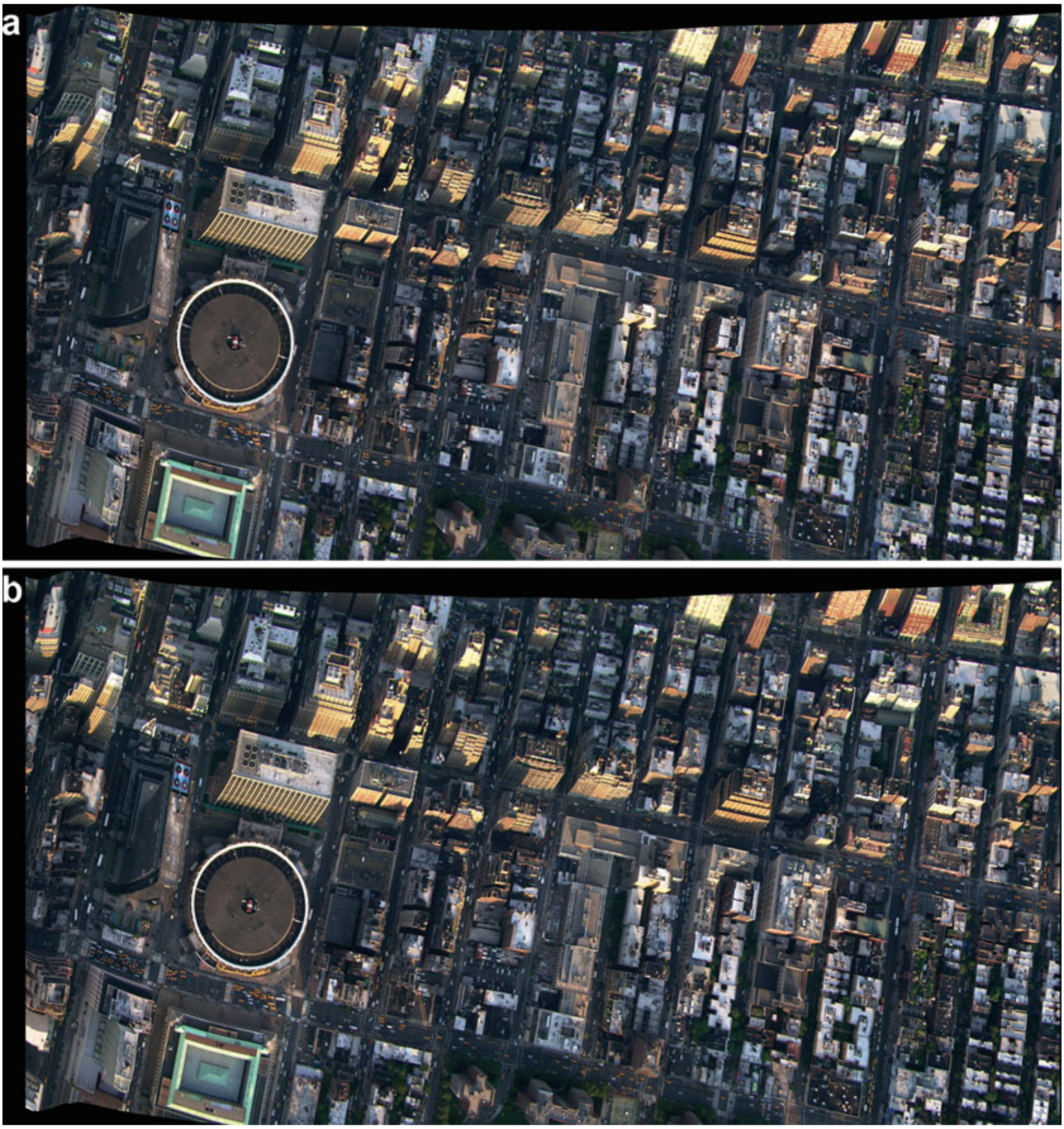
Video Mosaicing, Fig. 3 A pair of concentric mosaics of the City College of New York campus http://www-cs.engr.ccny.cuny.edu/~zhu/CSCI6716/CCNYCampus.jpg

Video Mosaicing, Fig. 4 A pair of pushbroom mosaics of the Amazon rainforest. http://www-cs.engr.ccny.cuny.edu/~zhu/57z10StereoColor.jpg

Rousso et al. 1998; Zhu et al. 2006), image-based rendering (Rademacher and Bishop 1998; Shum and Szeliski 1999, 2000; Zhu and Hanson 2006), aerial videography (Kumar et al. 1995; Peleg et al. 2000; Taylor and Andersen 2008; Zhu et al. 2004, 2005; Molina and Zhu 2014), document digitization (Iketani et al. 2006), microscopic imaging (Kose et al. 2017), multimodal alignment (Qu et al. 2010; Wang et al. 2013), and assistive vision (Colena et al. 2018).

In some of these applications, primarily 2D mosaics are used, assuming either the camera motion is (almost) a pure rotation or the scene is flat or very far from the camera, in order to avoid or reduce the motion parallax problem. When motion parallax cannot be avoided, 3D mosaics have to be considered. Methods have been proposed to generate mosaics, for example, for curved documents based on 3D reconstruction (Iketani et al. 2006), when the camera motion has translational components. Needless to say, with 3D reconstruction, a composite image with a new perspective view, or a new projection representation (such as orthogonal projection), can be synthesized from the original images. However, the drawback of this approach is that a full 3D reconstruction is needed, which is both computationally expensive and prone to errors. A more practical yet still fundamental approach without 3D reconstruction is to generate multi-perspective mosaics from a video sequence, such as mosaics on an adaptive manifold (Peleg et al. 2000), creating stitched images of scenes with parallax (Kumar et al. 1995) and creating multiple-center-of-projection images (Rademacher and Bishop 1998). When the dominant motion of the camera is translation, the projection model of the mosaic can be a parallel-perspective projection in that the projection in the direction of the motion is parallel, whereas the projection perpendicular to the motion remains perspective. This kind of mosaic is also called pushbroom mosaic (Tang and Zhu 2012) since the projection model of the mosaic in principle is the same as pushbroom imaging in remote sensing. A more interesting case is that by selecting different parts of individual frames, a pair of stereo mosaics can be generated that exhibit motion parallax, while each of them represents a particular viewing angle of parallel projection (Zhu et al. 2004; Tang and Zhu 2012). To generate stereo mosaics, the motion model is 3D, and therefore a bundle adjustment for 3D camera orientation is needed. The projection model is parallel-perspective, and therefore the composting surface is a plane that holds the parallel-perspective image. To generate a true parallel-perspective view in each mosaic for accurate 3D reconstruction, pixel selection is carried out for that particular viewing angle, and a coordinate transformation is performed based on matches between at least two original images for each pixel. A similar principle can be applied to concentric mosaics with circular projection (Peleg and Ben-Ezra 1999; Shum and Szeliski 1999).

In some applications such as surveillance and mapping, geo-referencing mosaicing is also an important topic. This is usually done when geo-location metadata is available, for example, from



Video Mosaicing, Fig. 5 A pair of pushbroom mosaics of a New York City scene

GPS and IMU measurements (Taylor and Andersen 2008; Zhu et al. 2005) taken with the video/images. Geo-referenced mosaics assign a geo-location to each pixel either by directly using the metadata from the video frames used to generate the mosaic or when metadata is not available; the video frames are aligned to a geo-referenced reference image such as a satellite image.

Video mosaicing techniques are also used for dynamic scenes, such as to generate dynamic pushbroom mosaics for moving target detection (Tang and Zhu 2012) and to create animated panoramic video textures in which different portions of a panoramic scene are animated with independently moving video loops (Agarwala et al. 2005; Rav-Acha et al. 2005).

Figure 2 shows a 360-degree panoramic mosaic represented on a cylindrical surface, which is generated from a video sequence taken by a video camera that roughly rotates around its optical center (Zhu et al. 2006). Figures 3 and 4 show two stereo mosaics that can be viewed with a pair of 3D glasses, red for the right eye and cyan for the left eye. High-resolution mosaics can be viewed by clicking the images in the figures in the online edition. The concentric stereo mosaic in Fig. 3 is generated from a video sequence taken by a handheld video camera that undertakes an off-center rotation with 360° of field of view coverage. Figure 4 is a pair of pushbroom stereo mosaics created from a video sequence taken by a camera looking down from an airplane flying over the Amazon rainforest (Zhu et al. 2004). Figure 5 (a) and (b) show a pair of stereo mosaics of a New York City scene from an airborne camera, in their original color format, from which a panoramic 3D map can be generated (Tang and Zhu 2012).

Future Directions for Research

Some open problems can be found in a good survey paper on image alignment and stitching (Szeliski 2006). These include robust alignments for stereo mosaics (or mosaics with motion parallax), mosaics for high dynamic range imaging and for super-resolution imaging and dynamic mosaics. Interesting areas in applications also include the use of drone video in aerial mapping and detection.

Cross-References

- ▶ Omnidirectional Vision
- ▶ Stereo Vision
- ▶ Visual Navigation

References

Agarwala A, Dontcheva M, Agrawala M, Drucker S, Colburn A, Curless B, Salesin D, Cohen M (2004) Interactive digital photomontage. ACM Trans Graph 23(3):292–300

Agarwala A, Zheng C, Pal C, Agrawala M, Cohen M, Curless B, Salesin D, Szeliski R (2005) Panoramic video textures. ACM Trans Graph 24(3):821–827

Burt PJ, Adelson EH (1983) A multiresolution spline with applications to image mosaics. ACM Trans Graph 2(4):217–236

Colena C, Kashyap N, Yau D, Cavalluzzi C, Zhu Z (2018) Panoramik: finding and locating objects via panoramic camera techniques. J Technol Pers Disabil, 6, 18–31,. CSUN Center on Disabilities

Colorado J, Mondragon I, Rodriguez J, Castiblanco C (2015) Geo-mapping and visual stitching to support landmine detection using a low-cost UAV. Int J Adv Robot Syst. https://doi.org/10.5772/61236

Dickson P, Li J, Zhu Z, Hanson A, Riseman E, Sabrin H, Schultz H, Whitten G (2002) Mosaic generation for under-vehicle inspection. In: IEEE workshop on applications of computer vision, Orlando, Dec 3–4

Eden A, Uyttendaele M, Szeliski R (2006) Seamless image stitching of scenes with large motions and exposure differences. In: IEEE computer society conference on computer vision and pattern recognition (CVPR'2006), 2498–2505

Harris C, and Stephens M (1988) A combined corner and edge detector. In: Fourth alvey vision conference, pp 147–151

Iketani A, Sato T, Ikeda S, Kanbara M, Nakajima N, Yokoya N (2006) Video mosaicing for curved documents based on structure from motion. ICPR 4:391–396

Irani M, Anandan P, Hsu SC (1995) Mosaic based representations of video sequences and their applications. In: ICCV, pp 605–611

Kose K et al (2017) Automated video-mosaicking approach for confocal microscopic imaging in vivo: an approach to address challenges in imaging living tissue and extend field of view. Sci Rep 7(10759)

Kumar R, Anandan P, Irani M, Bergen J, Hanna K (1995) Representation of scenes from collections of images. In: IEEE workshop on representations of visual scenes, pp 10–17

Lowe L (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110

Marzotto R, Fusiello A, Murino V (2004) High resolution video mosaicing with global alignment. CVPR 1:692–698

Molina E, Zhu Z (2014) Persistent aerial video registration and fast multi-view mosaicing. IEEE Trans Image Process 23(5):2184–2192

Peleg S, Ben-Ezra M (1999) Stereo panorama with a single camera. CVPR:1395–1401

Peleg S, Rousso B, Rav-Acha A, Zomet A (2000) Mosaicing on adaptive manifolds. IEEE Trans Pattern Anal Mach Intell 1144–1154

Qu Y, Khoo WL, Molina E, Zhu Z (2010) Multimodal 3D panoramic imaging using a precise rotating platform. In: IEEE/ASME international conference on advanced intelligent mechatronics (AIM 2010), Montreal, 6–9 July 2010

Rademacher P, Bishop G (1998) Multiple-center-of-projection images. In: Computer graphics proceedings, Annual conference series, pp 199–206

Rav-Acha A, Pritch Y, Lischinski D, Peleg S (2005) Dynamosaics: video mosaics with non-chronological time. In: IEEE computer society conference on computer vision and pattern recognition (CVPR'2005), pp 58–65

Rousso B, Peleg S, Finci I, Rav-Acha A (1998) Universal mosaicing using pipe projection. ICCV:945–952

Rzhanov Y (2013) Photo-mosaicing of images of pipe inner surface. SIViP 7(5):865–871. https://doi.org/10.1007/s11760-011-0275-z

Shum H-Y, Szeliski R (1999) Stereo reconstruction from multiperspective panoramas. In: Seventh international conference on computer vision (ICCV'99), pp 14–21

Shum H, Szeliski R (2000) Systems and experiment paper: construction of panoramic image mosaics with global and local alignment. Int J Comput Vis:101–130

Summan R, Dobie G, Guarato F, MacLeod C, Marshall S (2015) Image mosaicing for automated pipe scanning. AIP Conf Proc 1650:1334. https://doi.org/10.1063/1.4914747

Szeliski R (2006) Image alignment and stitching: a tutorial. Found Trends Comput Graph Vis 2(1):1–104

Tang H, Zhu Z (2012) Content-based 3d mosaics for representing videos of dynamic urban scenes. IEEE Trans Circuits Syst Video Technol 22(2):295–308

Taylor CN, Andersen ED (2008) An automatic system for creating geo-referenced mosaics from MAV video. In: IROS, pp 1248–1253

Wang T, Zhu Z, Taylor CN (2013) A multimodal temporal panorama approach for moving vehicle detection, reconstruction and classification. Comput Vis Image Underst 117(12):1724–1735. Special issue on machine vision beyond visible spectrum

Xu Y, Ou J, He H, Zhang X, Mills J (2016) Mosaicking of unmanned aerial vehicle imagery in the absence of camera poses. Remote Sens 8(3):204

Zheng JY, Tsuji S (1992) Panoramic representation for route recognition by a mobile robot. Int J Comput Vis 9(1):55–76

Zhu Z, Hanson A (2004) LAMP: 3D layered, adaptive-resolution and multi-perspective panorama – a new scene representation. Comput Vis Image Underst., Special issue on model-based and image-based 3D scene representation for interactive visualization, 96, 3, December, 294-326

Zhu Z, Hanson A (2006) Mosaic-based 3D scene representation and rendering. Signal Process Image Commun, 21, no 6, Elsevier B.V., October: 739-754. Special issue on interactive representation of still and dynamic scenes, Elsevier

Zhu Z, Hanson AR, Riseman EM (2004) Generalized parallel-perspective stereo mosaics from airborne video. IEEE Trans Pattern Anal Mach Intell:226–237

Zhu Z, Riseman EM, Hanson AR, Schultz HJ (2005) An efficient method for geo-referenced video mosaicing for environmental monitoring. Mach Vis Appl:203–216

Zhu Z, Xu G, Riseman E, Hanson A (2006) Fast construction of dynamic and multi-resolution 360° panoramas from video sequences. Image Vis Comput 24(1):13–26