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 mosaicking

Related Concepts

▸ Image 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, a higher dynamic range, or a better image resolution than any of the original images.

Background

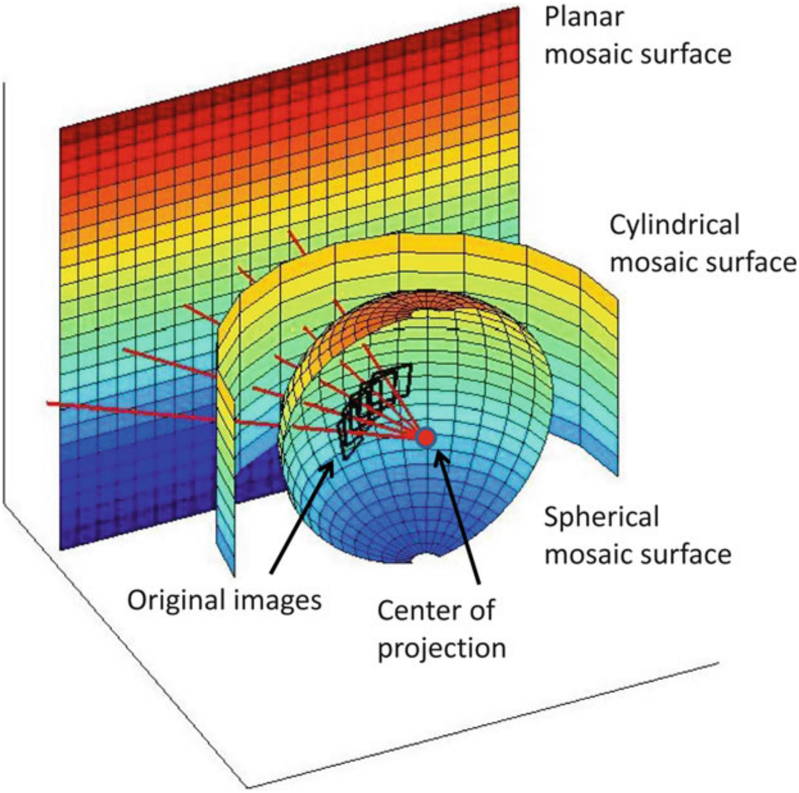
When collecting video of a scene or object, each individual image in the video may be limited compared to the desired final product, including limitations in the field of view, dynamic range, or image resolution. This is the case not only with personal video capture [1–3] but also with image-based rendering [4–6], aerial videography [7–11], and document digitization [12]. Generating mosaics with larger fields of view [2, 3, 5, 10, 12, 13], higher dynamic ranges [14], and/or higher image resolutions [15] facilitates video viewing, video understanding, video transmission, and archiving. When the major objective of video mosaicing is to generate a complete (e.g., 360°) 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 [2, 5, 6].

Theory and Application

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 the general principles can also be 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 certain amount of spatial

K. Ikeuchi (ed.), *Computer Vision*,
https://doi.org/10.1007/978-3-030-03243-2_68-1

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



overlap between two consecutive frames can also be viewed as a video sequence.

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 (with rotation, translation, scaling), an affine model, a perspective model (i.e., homography), or a full 3D motion model.

Many popular video mosaicing methods [16], for example, in [6, 14], 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 perspective projection (when FOV is less than 180°), a cylindrical projection (when FOV is 360° in one direction), or a spherical projection (when the FOV is full 360° in both directions). 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 applications such as surveillance, remote sensing, robot 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. In computer vision, motion parallax refers to the different changes in positions of images of objects at different distances caused by the translational 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 as de-ghosting [6], 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 [8, 11, 13]. 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 [3, 4, 7, 10], or a camera's optical center moving on a circular path [2, 5]. 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 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 and feature-based methods. In a direct method, 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. 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. But direct methods, especially those based on optical flow, can only be used when the inter-frame motion is relatively small. In a feature-based method, a feature detection operator such as the Harris corner or SIFT (scale invariant feature transform) detector is used first, and then the detected features are matched over the two frames to build up matches [16]. 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° circle in order to avoid gaps between the first and the last frame [6].

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 are discussed afterward.

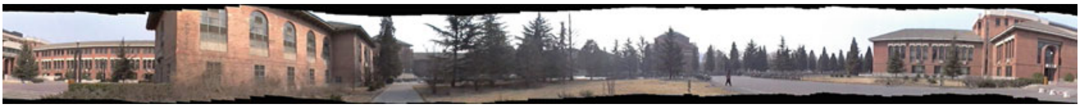
If the video sequence only has a few images, then one of the images can be selected as the reference image, and all the other images are warped and aligned with this reference image. In this case, the reference image with a perspective projection is the compositing surface, and therefore the final mosaic 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° 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 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 an original image frame (or a blending of multiple values from multiple original frames, see below), 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.

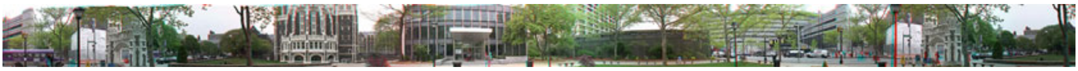
The third important issue in image composition is pixel selection and blending. Naturally in generating mosaics, there are overlaps among consecutive frames, resulting in two key questions: First, *where do we place the seam (i.e., the stitching line)* (the pixel selection problem)? Second, *how do we 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 [17] and gradient domain blending [1]. 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 [14].

Other benefits having multiple values from multiple images for each mosaiced pixel include high dynamic range imaging [14] and improved image resolution mosaicing [15]. 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.

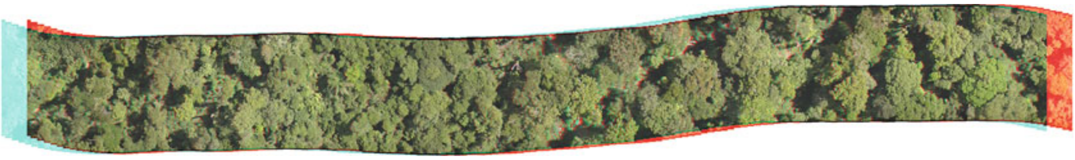
So far the discussions on image composition have focused primarily on 2D mosaics, assuming either the camera motion is (almost) a pure rotation or the scene is flat or very far from the camera, in order to avoid or reduce the motion parallax problem. When motion parallax cannot be avoided, 3D mosaics have to be considered. Methods have been proposed to generate mosaics, for example, for curved documents based on 3D reconstruction [12], 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 a full 3D reconstruction is needed, which is both computationally expensive and prone to noise. A more practical yet still fundamental approach without 3D reconstruction is to generate multi-perspective mosaics from a video sequence, under various names, such as mosaics on an adaptive manifold [8], creating stitched images of scenes with parallax [7] and creating multiple-center-of-projection images [4]. 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 push broom mosaic [18] since the projection model of the mosaic in principle is the same as push broom 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 represent a particular viewing angle of parallel projection [10]. To generate stereo mosaics, the



Video Mosaicing, Fig. 2 A 360° panoramic mosaic generated on a cylindrical surface



Video Mosaicing, Fig. 3 A pair of concentric mosaics of the City College of New York campus



Video Mosaicing, Fig. 4 A pair of push broom mosaics of the Amazon rain forest

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 [2,5].

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 GPS and IMU measurements [9, 11] 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, by aligning the video frames 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 push broom mosaics for moving target detection [18] and to create animated panoramic video textures in which different portions of a panoramic scene are animated with independently moving video loops [19, 20].

Open Problems

Some open problems can be found in a good survey paper on image alignment and stitching [16]. 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.

Experimental Results

Figure 2 shows a 360° 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. Figures 3 and 4 show two stereo mosaics that can be viewed with a pair of 3D glasses, red for the right eye and the 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 degrees of field of view coverage. Figure 4 is a pair of push broom stereo mosaics created from a video sequence taken by a camera looking down from an airplane flying over the Amazon rain forest.

References

1. 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
2. Peleg S, Ben-Ezra M (1999) Stereo panorama with a single camera. In: IEEE conference on computer vision pattern recognition (CVPR). IEEE Computer Society, Los Alamitos, pp 1395–1401
3. Rousso B, Peleg S, Finci I, Rav-Acha A (1998) Universal mosaicing using pipe projection. In: ICCV. Narosa Publishing House, New Delhi, pp 945–952
4. Rademacher P, Bishop G (1998) Multiple-center-of-projection images. In: Computer graphics proceedings. Annual conference series. Association for Computing Machinery, New York, pp 199–206
5. Shum H-Y, Szeliski R (1999) Stereo reconstruction from multiperspective panoramas. In: Seventh international conference on computer vision (ICCV'99), IEEE Computer Society, Los Alamitos, pp 14–21
6. Shum H, Szeliski R (2000) Systems and experiment paper: construction of panoramic image mosaics with global and local alignment. Int J Comput Vis 36:101–130
7. 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. IEEE Computer Society, Los Alamitos, pp 10–17
8. Peleg S, Rousso B, Rav-Acha A, Zomet A (2000) Mosaicing on adaptive manifolds. IEEE Trans Pattern Anal Mach Intell 22:1144–1154
9. Taylor CN, Andersen ED (2008) An automatic system for creating geo-referenced mosaics from MAV video. In IROS. IEEE, Piscataway, pp 1248–1253
10. Zhu Z, Hanson AR, Riseman EM (2004) Generalized parallel-perspective stereo mosaics from airborne video. IEEE Trans Pattern Anal Mach Intell 26:226–237
11. Zhu Z, Riseman EM, Hanson AR, Schultz HJ (2005) An efficient method for geo-referenced video mosaicing for environmental monitoring. Mach Vis Appl 16:203–216
12. Iketani A, Sato T, Ikeda S, Kanbara M, Nakajima N, Yokoya N (2006) Video mosaicing for curved documents based on structure from motion. In: ICPR, Hong Kong, vol 4, pp 391–396
13. Irani M, Anandan P, Hsu SC (1995) Mosaic based representations of video sequences and their applications. In: ICCV. IEEE Computer Society, Los Alamitos, pp 605–611
14. 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). IEEE Computer Society, Los Alamitos, pp 2498–2505
15. Marzotto R, Fusiello A, Murino V (2004) High resolution video mosaicing with global alignment. In: CVPR, vol 1. IEEE Computer Society, Los Alamitos, pp 692–698
16. Szeliski R (2006) Image alignment and stitching: a tutorial. Found Trends Comput Graph Vis 2(1):1–104
17. Burt PJ, Adelson EH (1983) A multiresolution spline with applications to image mosaics. ACM Trans Graph 2(4): 217–236
18. Tang H, Zhu Z, Wolberg G (2006) Dynamic 3D urban scene modeling using multiple pushbroom mosaics. In: 3DPVT, Chapel Hill, USA, pp 456–463
19. 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
20. 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). IEEE Computer Society, Los Alamitos, pp 58–65