A Methodical Review on Image Stitching and Video Stitching Techniques

S. Pravenaa
Research Scholar, School of Electronics Engineering,
VIT University, Chennai-600 127 Tamil Nadu, India.

Dr. R. Menaka
Associate Professor, School of Electronics Engineering,
VIT University, Chennai-600 127 Tamil Nadu, India.

Abstract
Image stitching is a technique in which several picturesque images of overlapping domain of view are blended together to result in a panoramic image of high resolution. But most methods of image stitching require almost precise intersects between images and identical illumination to get picture perfect outcomes. Image stitching surveys depict that image stitching is till now a perplexing issue for panoramic images. Image stitching and Video stitching are the current research area in the fields of computer vision, computer graphics and Photographics. [1] This article outlines the aspects of Image and Video stitching techniques, different process stages and approaches adapted, along with different views of map projection.

Keywords: Image stitching, Video stitching, Registration, Calibration, Blending, Map projection

Introduction
Image stitching finds applications in Video stabilization, Video stitching, Medical Imaging, Object Insertion, Summarization, Texture synthesis, High resolution photo mosaics etc.. The terminologies used in Image Stitching are as follows:

Feature
A feature is a substantial evidence mined out of an image to provide an exhaustive description of the image. To be more precise, features are entities, that are unique, prominent and desirable, extended over the image and easy to perceive. The objects can be segmented areas, intensity etc...

Image registration:
A sample space of data obtained by sampling the similar view or object at various time intervals, or from altered viewpoints, will result in diverse coordinate systems. Image registration is a procedure in which transformation of different data sets into a single coordinate system is achieved. Registration is essential to associate or incorporate the data obtained from unlike dimensions.

Feature Matching:
The intention of feature matching is to create similarities between two feature sets. The bottleneck can be an inappropriate detection of feature or distortion in image. A matching algorithm need to be devised to deal with single features i. e. feature that is only seen in one image.

Image Stitching Approach
Before Image registration and alignment, there is a need to establish a mathematical relationship between pixel coordinates of one image to another. An assortment of parametric motion models are achieved from 2D transforms, 3D camera rotations, planar perspective models, lens distortions, and mapping to non-planar surfaces. Image registration methods come in to estimate its parameters, only when an appropriate motion model is finalized to define the orientation between image pairs. Image stitching techniques are broadly classified into two approaches: Direct and Feature based techniques. [2] Direct techniques compare every pixel’s intensity of an image with that of the other, whereas Feature based techniques extract distinct features from the processed images and correlate them. This approach is more advantageous in being more robust against scene rotation, faster, and has the proficiency to automatically determine the overlapping relationships among an unordered set of images.

In direct technique, pixel to pixel variation is reduced to do image stitching. While feature-based technique works by pulling out a feature sets and then does the matching with each other. [3]

Direct (Pixel-Based) techniques
In direct technique, each pixel’s intensity of image is compared with the other. The main benefit of direct technique is that, it lessens the sum of absolute differences between overlying pixels. In this technique, each pixels compared with each other so it’s a very complex technique. They are variant to image scale and rotation. Direct method optimally uses the information obtained from the image alignment. It measures the role of each and every picture element in the image. The main limitation of this technique is limited range of convergence. Direct Method uses information from all pixels. Approximation of homography [4] is always updated to minimize the cost function. Some parameters of the homography are predicted with the help of Phase-Correlation. Image alignment and Image stitching essentially need to determine the similar pixels in the overlapping area between the images. Certain transformation happens so that, there exists a need to remap pixel coordinates from one image into
another, when a same camera is used to take two consecutive photos with a slight shift while capturing the second one. The factors of these transformations can be grouped in a Homography matrix. The intricacy of the matrix corresponds to the various classes of transformation.

In this approach the images are shifted or warped with respect to each other to find the degree of similarity between pixels. Such methodologies which use pixel-to-pixel matching are commonly known as direct methods. For this, suitable error metric is preferred to compare the images. There also Fourier based and Incremental algorithms for direct based matching. However, all these methods require computation for each pixel. This comes out to be computationally intensive if the complexity of algorithm is high. Hence these methods are usually avoided in commercial software for image stitching.

Feature-based techniques
In feature-based technique, all main feature points in an image pair is compared with that of every feature in other image by making use of local descriptors. For feature-based image stitching techniques-feature extraction, Image registration, and Image blending are the various stages followed. Feature-based methods are used by instituting equivalences between points, lines, edges, corners or other shapes. The main uniqueness of robust detectors incorporates invariance to noisy image, scale invariance, translation invariance, and rotation transformations.

There are numerous feature detector methods exist, out of which some are, Harris, [5], [6] Scale-Invariant Feature Transform (SIFT), [7], [8] Speeded Up Robust Features (SURF), [9], [10] Features from Accelerated Segment Test (FAST) [11], PCA-SIFT12] and ORB[13] techniques. For any category of movement of scene happened in image, feature based technique is advantageous because of its robustness. This method is very faster and it has the capability to identify panoramas by automatically detecting the adjacency relationship between an ordered image set. These features are best matched for completely automated panoramas stitching. Feature based techniques rely on precise recognition of image features. Correlation of features leads to computation of the camera motion that can be tested for alignment.

Video Stitching Approach

Process Stages
Image Stitching Stages
The image stitching can be performed in three main steps:
- Image calibration
- Image registration
- Image blending

Image calibration produces an estimate of the intrinsic and extrinsic camera parameters. In image registration, given set of input images are evaluated to find the translations, used for the arrangement of images. These images are combined together to form a distinct image, after registration. In the following subsections, these main steps are discussed briefly.

![Figure 1: Main Steps of Image Stitching](image-url)

**Calibration**
Image calibration lessens the dissimilarities between an idyllic lens model and the combination of camera-lens used. These differences result from optical defects such as deformations and exposure differentiation between images. [16] Intrinsic and extrinsic camera parameters are improved in order to reconstruct the three-dimensional structure of a pixel coordinates scene of its image points. Intrinsic camera parameters associate the pixel coordinates of a particular image with the equivalent coordinates in the camera reference frame. Extrinsic camera parameters characterize the position and direction of the camera reference frame in accordance with a known reference frame.

**Image Registration**
Image registration is defined as the method of lining up more than two images which are captured from different perspective points. The objective of image registration is to construct geometric association between images. Therefore, images can be compared and applied to subsequent steps appropriately. [17] Image blending is processed to obtain a smoother transition from one image to another image, in order to get a seamless image. [18]

Image registration comprises many different methodologies depending on dimensionality, feature types etc.. Most of the feature based registration techniques can be classified into four steps.
1. Feature detection-relevant and unique features are identified.
2. Feature matching-feature correspondences between the images are established.
3. Transform model-feature correspondences that align the images for the transformation are determined by the use of the correspondences.
4. Image transformation-images are aligned.

**Image Blending**
Image blending is employed across the stitch, in order to result in seamless stitching. There exists two approaches of
blending the images. [19] One is the alpha “feathering” blending and the other one is Gaussian pyramid. Alpha feathering blending considers weighted average of two images. Alpha blending is best suited for well aligned image pixels and the main difference between the images is only the overall shift in intensity. Gaussian pyramid approach amalgamates the images at diverse spectrum and sort out them accordingly. The lower the spectrum, the more the boundary is blurred. Gaussian pyramid results in blurring of the boundary meanwhile maintaining the pixels away from the boundary.

Video Stitching Stages

The process of video stitching is broken down into 3 parts:[20]

Registration
This stage involves finding and matching the features from individual video feeds.

Calibration
This stage aims to correct distortions, exposure differences, vignetting (reduction of brightness of image and/or saturation at the fringe of the image) and chromatic aberrations.

Blending
This stage involves stitching the individual feeds together as well as executing the adjustments found in stage 2. Often, stage 2 and stage 3 are combined into one stage. This stage also aims to minimize the visibility of the stitch seams. [21]

Below are some of the elementary models, used to conclude vital information in image stitching approaches. [22]

Harris Corner Detection
Corners are nothing but areas of interest in the image with great deviation in intensity in all the directions.
The mathematical form basically finds the difference in intensity for a displacement of \((u, v)\) in all directions. This is expressed as below:

\[
E(u, v) = \sum_{x, y} w(x, y) \left( \frac{I(x + u, y + v) - I(x, y)}{\text{intensity}} \right)^2
\]

Window function is either a rectangular window or gaussian window which gives weights to pixels underneath. The function \(E(u, v)\) is to be maximized for corner detection.

Feature point Detection

Required characteristic of feature point detectors for image stitching are:

1. A huge sample of feature points—a sufficient number of feature points to estimate the transformation matrix.
2. Transformation invariance—feature points should be invariant to image transformations.
4. Efficient in complexity.

Below are some of the elementary models, used to conclude vital information in image stitching approaches. [22]

Harris Corner Detection
Corners are nothing but areas of interest in the image with great deviation in intensity in all the directions.
The mathematical form basically finds the difference in intensity for a displacement of \((u, v)\) in all directions. This is expressed as below:

\[
E(u, v) = \sum_{x, y} w(x, y) \left( \frac{I(x + u, y + v) - I(x, y)}{\text{intensity}} \right)^2
\]

Window function is either a rectangular window or gaussian window which gives weights to pixels underneath. The function \(E(u, v)\) is to be maximized for corner detection.
When an image is scaled, a corner may not look like a corner. When zoomed in the same window, a corner in an image inside a small window is almost flat. Hence Harris corner is not scale invariant. Thus evolved a new algorithm, Scale Invariant Feature Transform (SIFT). [23] SIFT [24] has been a standard algorithm in image stitching, and it has proved with significant matching performance. But the algorithm levies a massive computational encumbrance in locating key point and image-scale space description, which makes it unsuitable for real-time video stitching.

This algorithm is based on SIFT [25] and recommends an essential parameter for estimating the performance of affine recognition (transition tilt) [26]. The transition tilt processes the amount of change in viewpoint from one view to another, thus providing an initial instinctive method to absolute tilt and transition tilt. It clarifies the need of necessity to find a fully affine invariant recognition by simulating large tilts on matched images. In practical, transition tilts are much bigger than absolute tilts. Even, transition tilts can be the square of absolute tilts.

SURF
SURF is inspired by SIFT algorithm. Integral image is made use in the computation of approximate determinant of Hessian blob detector. Further, integral image is adopted to compute the sum of Haarwavelet[27] response as descriptor around the point of interest. SURF is many fold times faster than SIFT. Hessian-matrix approximation is used for interest point detection, because of its consistent accuracy. In addition to it, blob-like structures are detected at locations where the determinant is optimal. SURF and SIFT [28] are similar with respect to image feature detection steps, but SURF is more efficient in reducing the computational time of key-points with the use of integral images. [29]

ORB ( Oriented FAST and Rotated BRIEF)
ORB [30], built on FAST key point detector and BRIEF descriptor [31], is used real-time extraction of key point features. FAST and its variants are much efficient in finding reasonable key points. Also, this technique finds application in many real-time systems such as mobile augmented reality, parallel tracking and mapping. [32], [33]

FAST makes use of a single parameter, namely the intensity threshold between center pixel and its adjacent pixels on a circular ring. FAST has a constraint of non-inclusion of an orientation operator such as the histograms of the gradient in SIFT [34] and SURF. FAST technique’s principle of giving large responses along edges also serves as a bottleneck. BRIEF is a technique which uses bit string descriptor, which is derived from an image patch that is created from a sample space of binary intensity tests. Hamming distance, which is very effective in calculating a binary string is made use in the evaluation of ORB descriptor. BRIEF has a limitation is that it is not invariant to rotation.

In order to manage the rotation problem, ORB makes use of the intensity centroid to compute a direction for every FAST corner, after which Harris corner is implemented to obtain a response to eliminate the points alongside the edges.

General overview of Image stitching algorithms
To relate pixel coordinates in one image to another, mathematical model has to be developed. Such models are developed using algorithms for image alignment. Next arises a need for evaluation of the accurate alignments of various pairs of images. [35] Algorithms that coalesce direct pixel-to-pixel evaluation with gradient descent and further optimization techniques shall be made use inorder for parameters estimation. Correspondences between pairs of images is obtained by efficiently matching distinctive features in each image. For computing worldwide constant set of alignments and to effectively find out the images which overlap with one another, several techniques have been improved for multiple images which exist in a panorama. [36]

General overview of Video stitching algorithms
Image registration is the first step in video stitching. Lowe’s SIFT( Scale Invariant Feature Transform)[37] is the best suited algorithm in recognizing interest points, as these points are invariant to orientation, scale, distortion and noise. The second step in video stitching is calibration. Images are added to the panorama, one at a time, which are then aligned to the images currently in the panorama already. The last step in video stitching is blending, in which the video streams are stitched together to create one panoramic video stream. [38] In this stage, all visual artifacts are minimized, so as to give an impression that the panoramic video was recorded using one camera.

Possible issues with Image stitching and Video stitching techniques
Stitching two images may result in a visible seam, since the background illumination in two views cannot be assured to be the same. For the same continuous foreground, seam will result if there is any change in background between two images. [39]
The key concerns to address are existence of parallax, scene motion, exposure differences and lens distortion. The best image set should have a rational extent of overlap of at least 15-30%, for panoramic stitching to surmount lens distortion and sufficient measurable features. To minimize the probability of occurrence of seam, the images set should have consistent exposure between frames. But in non ideal case, the contrast and intensity through the frames vary as the intensity varies across the whole scene. As a result, ghosting occurs due to Lens distortion, gesture in the scene and misalignment. And the ratio of width to height of panorama image should be considered to create a visually pleasing panorama.

Video stitching experiences many challenges compared to image stitching, such as temporal coherence, view on foreground moving objects and camera jittering. [40]

Projective layouts
Various map projections can be used to arrange stitched images, for image segments taken from same point in space.
Rectilinear
In Rectilinear projection, two-dimensional plane view of stitched image is possible, when all panosphere converges in a single point. In spite of image directions, lines that seem to be straight are shown as straight. Rigorous distortion is found near the image boundary when viewed around 120° or other. In panorama viewing, exercising of cube faces with cubic mapping is found to be a major purpose of rectilinear projection. [41]

Cylindrical
In Cylindrical projection, stitched image is shown as a 360° horizontal field of view with a restricted vertical field of view. Cylindrically projected Panoramas are looked as if the image is draped into a cylinder and observed from inside. If observed on a two-dimensional plane, horizontal lines are shown as curved lines and vertical lines are shown as straight lines. When nearing the top of the panosphere, vertical distortion gets increased rapidly. [41]

Spherical
A derivative of cylindrical projection is Spherical projection or equirectangular projection. In this projection, the stitched image is shown as a 360° horizontal by 180° vertical field of view resulting in a complete sphere. Spherically projected panoramas are looked as if the image is enfolded into a sphere and seen from inside. If viewed on a two-dimensional plane, horizontal lines are shown as curved lines and vertical lines are shown as vertical lines itself. [41]

Stereographic
To set the field of view to show the entire ground and areas above it, the virtual cameras are fixed straight down. To create little planet panorama, Stereographic projection or fisheye projection is made use. Tunnel effect is created by pointing the virtual camera upwards. Visually pleasing result is produced due to conformality of stereographic projection. [41]

Discussion and Conclusion
This paper discusses the three main steps of image stitching, namely calibration, registration and blending. Also, two main approaches of image stitching namely, direct techniques and feature-based techniques are enumerated. Video stitching techniques such as Registration, calibration and Blending are also highlighted in this paper. In general, Image registration’s objective is to find the geometric correlation between the input images, while image blending is responsible for building a seamless constitution. Apart from the algorithms carried out in the above techniques, an insight of different map projections are also dealt in this paper.

References


© Research India Publications. http://www.ripublication.com
