Available at: www.researchpublications.org

Image Stitching Techniques-An overview

Ms. Vrushali S. Sakharkar
Computer Science and Engineering
PRMIT & R, Badnera, Amravati
vrushu.sakharkar@gmail.com

Prof. Dr. S. R. Gupta
Computer Science and Engineering
PRMIT & R, Badnera, Amravati

Abstract — **In this paper, we describe various Image stitching techniques used to combine multiple images together to make a wide angle picture called Panorama. Algorithms for stitching and aligning have many applications in computer vision like texture synthesis, object recognition, medical imaging. Images of any object or place are acquired using 360· rotating camera and the overlapped portion of images should be aligned properly. Image stitching requires identical exposures to produce seamless results. It is also known as mosaicing. Stitching can be done into three stages Acquisition, Remapping, Blending. We have described different techniques for all the stages. Image acquisition can be done using rotating camera, translating camera and hand held camera. Image Remapping involves Alignment, conversion of camera image to Sphere, Selection of final projection, Exposure setting. Image remapping is also called Warping. The final warped images should blend perfectly so the visible seams between the images can be removed. Image blending techniques like Average Blending, Alpha Blending, Feathering, Pyramid blending used to remove Exposure difference, Blurring, Ghosting. Finally unwanted information can be cropped using various cropping techniques like graph cuts, gradient blending etc.**

Keywords-component: **Image Stitching, Image Blending, Image Warping.**

I. INTRODUCTION

Image stitching or photo stitching is the process of combining multiple photographic images with overlapping fields of view to produce a segmented panorama or high-resolution image. Commonly performed through the use of computer software, most approaches to image stitching require nearly exact overlaps between images and identical exposures to produce seamless results. It is also known as mosaicing. "Stitching" refers to the technique of using a computer to merge images together to create a large image, preferably without it being at all noticeable that the generated image has been created by computer. Algorithms for aligning images and stitching them into seamless photo-mosaics are among the oldest and most widely used in computer vision. Frame-rate image alignment is used in every camcorder that has an "image stabilization" feature. Image stitching algorithms create the high resolution photo-mosaics used to produce today's digital maps and satellite photos. They also come bundled with most digital cameras currently being sold, and can be used to create beautiful ultra wide-angle panoramas.
We describe various Image Stitching phases in section 1.In section 2,we give details of Image Acquisition phase. Image remapping is an essential process of Image Stitching described in section3. Finally aligned images are blended together to avoid seams in the stitched image this procedure is known as Image Blending specified in Section 4.

A. Different kinds of Image Stitching

There are different kinds of Stitching,

- Mosaic - stitch multiple rows of pictures that were taken without rotating the camera around a single point, but with the camera kept perpendicular with the subject.
- Panorama (single-row) - stitch a single row of pictures (created by rotating the camera around a single point in a flat plane, which is normally parallel with the horizon).
- Panorama (multi-row) - stitch multiple rows of pictures (created by rotating the camera around a single point in a flat plane but tilting or pitching the camera up and/or down so that for each row of pictures the lens is not necessarily parallel with the plane of rotation).
- Panorama (pano-camera) - just stitch together the ends of panoramic picture created with a panoramic camera.
- Spherical panorama - stitch any number of pictures in such a way as to create a spherical panorama, the important distinction between this and single or multi-row panoramas is that the "poles" (i.e the very top and bottom of the image) must be stitched also so that the user can look straight up and down and see a smoothly blended image.

1) Stages of the Image Stitching Process

a) Image Acquisition:

This stage of Image stitching is to do calibration of Images. It requires selection of the position and acquisition of images. In this step, a decision needs to be made on the type of resultant panoramic images. According to the required panoramic images, different image acquisition methods are used to acquire the series of images.

b) Image Remapping (Image Warping):

Remapping is the process of changing the image geometry in order to fit to the adjacent images that should contribute to a panorama. This remapping process "warps" the images in such a way so that they can be aligned perfectly. Without this remapping process, the images will not join together correctly. It includes conversion of camera image to sphere, Image alignment, setting screen projection.

c) Image Blending:

Once the source pixels have been remapped onto the final composite surface, we must still decide how to blend them in order to create an attractive looking panorama. Blending is used to remove visible seams (due to exposure differences), blurring (due to miss-registration), or ghosting (due to moving objects).

d) Cropping:

Image cropping refers to removing unwanted areas from a photographic or illustrated image. One of the most basic Image manipulation processes, it is performed in order to remove an unwanted subject or change its aspect ratio, or to improve the overall composition.

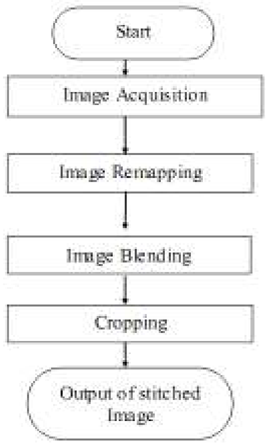


Fig.1 Flowchart of Producing Panoramic Image

II. IMAGE ACQUISITION

Different image acquisition methods can be used to acquire input images that produce different types of panoramic images, depending on the type of panoramic images required and the availability of equipment. Three set-up's to acquire images for panoramic image generation are described and discussed in this section. In the first set-up, the camera is set upon a tripod and the images are obtained by rotating the camera. The second set-up places the camera on a sliding plate and the images are obtained by shifting the camera on the sliding plate. The third set-up is where the camera is held in a person's hands and the person takes the images by turning around on the same spot, or walking in a direction perpendicular to the camera's view direction. In all three set-up's, a still image camera has been used to take the images. The camera co-ordinate system is shown in Fig,2 where the Z-axis points towards the object of interest and the Y-axis passes through the optical axis of the camera.

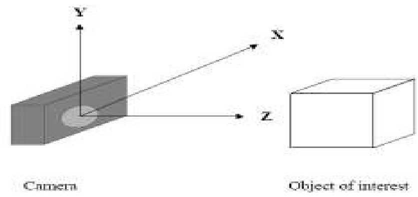


Fig. 2 Camera co-ordinate system

The camera's angles of view in the horizontal and vertical directions decide each image's coverage in the horizontal and vertical directions. The angles of view are defined in Fig.3, where the angles and respectively represent the camera's angles of view in the horizontal and the vertical directions.

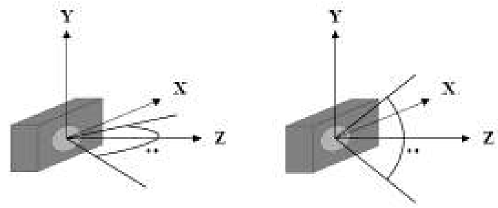


Fig. 3 The horizontal and vertical angles of view.

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

A. Acquisition by Camera Translations:

In this acquisition method, the camera is shifted in a direction parallel to the imaging plane. On a smaller scale, the camera can be mounted onto a sliding plate to achieve the translations. The camera and the sliding plate are placed directly in front of the objects of interest and an image is taken with each translation of the camera until the series of images cover the desired range. Fig 4. Shows the setup of this method, where the camera is aligned with the sliding plate so that the imaging plane parallel to the orientation of the sliding plate is.

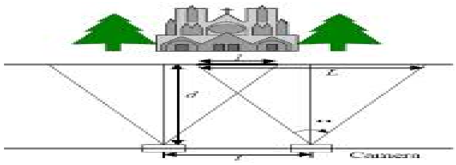


Fig.4 Geometry for image acquisition by camera translations.

Given the camera translation, *t*, the distance between camera and object of interest, *d*, and the camera's horizontal angle of view, the ratio of the overlapping region to the whole image can be estimated by Eqn. 2,

$$l / L = 1 - t / 2d \tan(1/2) \quad ---- \quad (2)$$

Nevertheless, the actual size of the overlapping region between successive images is determined by the accuracy in setting up the camera. In acquiring images by translation, it is important to ensure that the image planes are parallel to the direction of camera translations. Otherwise, the size of the objects in the images varies as the camera is shifted, causing problems in image stitching. One disadvantage of this method is that the required translation, t, increases as the distance between the camera and the object of interest, d, increases, if the acquired images are to have the same sized overlapping region. Therefore, acquiring images when the object of interest is far away from the camera is more difficult due to the magnitude of required translations. Furthermore, since the acquired images are all on the same plane, the panoramic images obtained from images acquired by translation does not provide the feeling of looking into a 3D environment as in the case for the panoramic images obtained from images acquired by rotation.

B. *Acquisition by a Hand held Camera:*

This acquisition method is comparatively easy to perform; the user simply holds the camera and takes images of his/her surroundings by either rotating on the same spot or moving in a predefined direction roughly parallel to the image plane. However, images acquired by this method can be difficult to stitch, due to the amount of unwanted camera rotation or translation during the acquisition of images.

In the case of the user turning with the camera to obtain the images, the user acts as the tripod of the camera in Section 2.3.1. However, rather than rotating about, or approximately about the vertical axis of the camera, there are inaccuracies in the alignment of the rotating axis with the vertical axis. It is also difficult to control the angles rotated between successive images. Therefore, the sizes of the overlapping regions between adjacent images have a greater variation than for images acquired by a camera mounted on a tripod.

When the user holds up the camera and moves in one direction to acquire the images, the user acts as the sliding plate in acquisition by translation in Section 2.3.2. However, in this situation, it is even more difficult to control the distance shifted between each image and keep each image on the same imaging plane. Therefore, apart from the difference in the size of the overlapping regions, the image planes of the acquired images have different orientations and cause problems in image stitching. It is often desirable to have larger overlapping regions between adjacent images to reduce the effects of the above mentioned problems in acquiring images free-handed. Larger overlapping regions implies that the camera rotations, or translations between successive images are smaller, thus reducing the amount of inconsistencies between images. Nevertheless, acquiring images by a hand held camera is very easy to manage and can be performed in many locations where it might be difficult to set up equipment such as the tripod or the sliding plate. If care is taken during the acquisition of the images, it is possible to produce panoramic images of similar quality to those generated with images acquired by mounted cameras.

III. IMAGE REMAPPING (IMAGE WARPING)

Remapping is the process of changing the image geometry in order to fit to the adjacent images that should contribute to a panorama. This remapping process "warps" the images in such a way so that they can be aligned perfectly. Without this remapping process, the images will not join together correctly. It highly depends on aligning, since any alignment change of an image requires a different remapping. Remapping is an inherent part of the stitching process. Remapping includes the following steps

A. *Conversion of camera image to sphere:*

The camera image has to be transformed into angle space, which here is visualized as projecting the image from a virtual camera, reversing the light path, onto a sphere

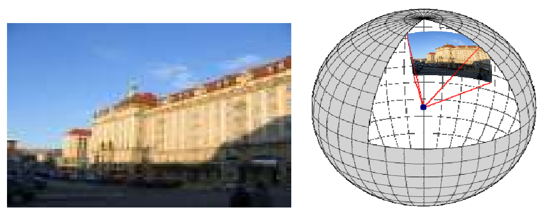


Fig. 5Conversion of camera image to sphere

B. *Image Alignment (Image Registration):*

Generally, one is free to choose any orientation of the virtual camera/projector. But for panoramic images, one wants to align these images. The usual software approach is to mark a set of same features in image pairs, and have the distance between such markers pairs minimized in an overall optimization process. In this section we describe different image alignment techniques.

Available at: www.researchpublications.org

Fig. 6 Image Alignment by matching features

1) Feature based registration:

A feature (key point) is a significant piece of information extracted from an image which provides more detailed understanding of the image. In other words, features are objects, which should be distinct, salient and preferable frequently spread over the image and easy to detect. The objects can be for example landmarks or segmented areas and even intensity can be seen as a feature. The number of common detected objects (seen in both images) must be sufficiently high to be able to perform a good estimation of the transformation model. The features can then be seen as points, lines, and regions. The first task of the registration is to decide what kind of feature we search for, and how to represent them. Most of the feature based registrations techniques can be divided into four steps.

a) Feature detection:

The salient and distinctive features are detected. After deciding what kind of features we are looking after we apply suitable detector method on the images. The method could be either worked interactively with an expert or automatically. The choice is depending on the situation. Automatically methods usually applied when extrinsic methods are used, or when geometrically landmarks are used. For region features we can use segmentation methods. Line features are usually detected by means of a edge detector like the Canny detector. When we are looking for point features corner detectors e.g. the Harris corner detector is suitable.

b) Scale-invariant feature transform (SIFT):

It is an algorithm to detect and describe local features in images The SIFT features are local and based on the appearance of the object at particular interest points, and are invariant to image scale and rotation. They are also robust to changes in illumination, noise, and minor changes in viewpoint. In addition to these properties, they are highly distinctive, relatively easy to extract, allow for correct object identification with low probability of mismatch and are easy to match against a (large) database of local features. Object description by set of SIFT features is also robust to partial occlusion; as few as 3 SIFT features from an object are enough to compute its location and pose.

c) *Holography*:

Holography between pairs of images is then computed using RANSAC and a probabilistic model is used for verification. Because there is no restriction on the input images, graph search is applied to find connected components of image matches such that each connected component will correspond to a panorama. Finally for each connected component Bundle adjustment is performed to solve for joint camera parameters, and the panorama is rendered using multi-band blending.

2) Recognizing panorama:

To perform fully automated image stitching is a technique to recognize which images actually go together, which Brown and Lowe (2003) call *recognizing panoramas*. If the user takes images in sequence so that each image overlaps its predecessor and also specifies the first and last images to be stitched, bundle adjustment combined with the process of *topology inference* can be used to automatically assemble a panorama (Sawhney and Kumar 1999). However, users often jump around when taking panoramas, e.g., they may start a new row on top of a previous one, or jump back to take a repeated shot, or create 360° panoramas where end-to-end overlaps need to be discovered. Furthermore, the ability to discover multiple panoramas taken by a user over an extended period of time can be a big convenience.

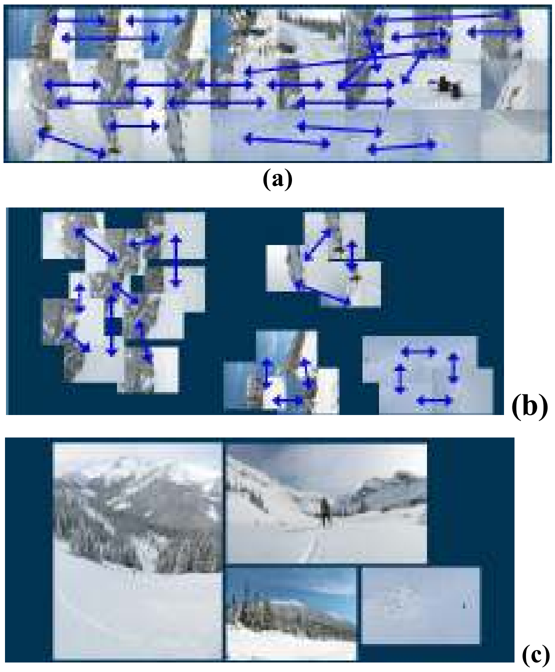


Fig. 7 Recognizing panoramas (Brown et al. 2004): (a) input images with pair wise matches; (b) images grouped into connected components (panoramas); (c) individual panoramas registered and blended into stitched composites.

To recognize panoramas, Brown and Lowe (2003) first find all pair wise image overlaps using a feature-based method and then find connected components in the overlap graph to “recognize” individual panoramas (Figure 8). The feature-based matching stage first extracts SIFT feature locations and feature descriptors (Lowe 2004) from all the input images and then places these in an indexing structure, as described in §4.2. For each image pair under consideration, the nearest matching neighbor is found for each feature in the first image, using the indexing structure to rapidly find candidates, and then comparing feature descriptors to find the best match. RANSAC is then used to find a set of *inlier* matches, using a pairs of matches to hypothesize a similarity motion model that is then used to count the number of inliers.

C.Selection of Projection:

The first choice to be made is how to represent the final image. If only a few images are stitched together, a natural approach is to select one of the images as the *reference* and to then warp all of the other images into the reference coordinate system. The resulting composite is sometimes called a *flat* panorama, since the projection onto the final surface is still a perspective projection. For larger fields of view, however, we cannot maintain a flat representation without excessively stretching pixels near the border of the image. (In practice, flat panoramas start to look severely distorted once the field of view exceeds 90˚ or so.) The usual choice for compositing larger panoramas is to use a cylindrical (Szeliski 1994, Chen 1995) or spherical (Szeliski and Shum 1997) projection. In fact, any surface used for *environment mapping* in computer graphics can be used, including a *cube map* that represents the full viewing sphere.

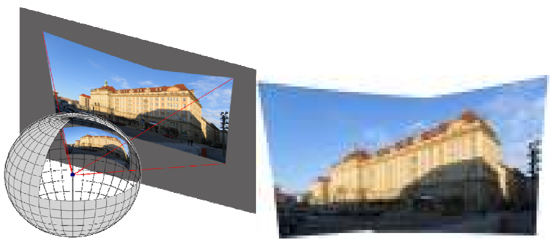


Fig. 8 Projection of sphere image to screen

D.Selection of View

Once we have chosen the output parameterization, we still need to determine which part of the scene will be *centered* in the final view. As mentioned above, for a flat composite, we can choose one of the images as a reference. Often, a reasonable choice is the one that is geometrically most central. For example, for rotational panoramas represented as a collection of 3D rotation matrices, we can choose the image whose z-axis is closest to the average z-axis (assuming a reasonable field of view). Alternatively, we can use the average z-axis (or quaternion, but this is trickier) to define the reference rotation matrix.

For larger (e.g., cylindrical or spherical) panoramas, we can still use the same heuristic if a subset of the viewing sphere has been imaged. If the case of full 360◦ panoramas, a better choice might be to choose the middle image from the sequence of inputs, or sometimes the first image, assuming this contains the object of greatest interest. In all of these cases, having the user control the final view is often highly desirable.

E.Exposure Setting:

Input images need not have the same exposure, or white balance. Together with vignetting (images are darker in the corners than in the middle) the required corrections can be derived from images with a suitably large overlap area.

IV. IMAGE BLENDING

Once we have registered all of the input images with respect to each other, we need to decide how to produce the final stitched (mosaic) image. This involves selecting a final compositing surface (flat, cylindrical, spherical, etc.) and view (reference image). It also involves selecting which pixels contribute to the final composite and how to optimally blend these pixels to minimize visible seams, blur, and ghosting. Once the source pixels have been mapped onto the final composite surface, we must still decide how to blend them in order to create an attractive looking panorama. If all of the images are in perfect registration and identically exposed, this is an easy problem (any pixel or combination will do). However, for real images, visible seams (due to exposure differences), blurring (due to mis-registration), or ghosting (due to moving objects) can occur. Creating clean, pleasing looking panoramas involves both deciding which pixels to use and how to weight or blend them. In this section we describe different image blending techniques.

A. Alpha blending:

1) Alpha Channel:

The alpha channel is normally used as an opacity channel. If a pixel has a value of 0% in its alpha channel, it is fully transparent (and, thus, invisible), whereas a value of 100% in the alpha channel gives a fully opaque pixel (traditional digital images). Values between 0% and 100% make it possible for pixels to show through a background like a glass, an effect not possible with simple binary (transparent or opaque) transparency. It allows easy image compositing. Alpha channel values can be expressed as a percentage, integer, or real number between 0 and 1 like RGB parameters. In graphics, a portion of each pixel's data that is reserved for transparency information. 32-bit graphics systems contain four channels -- three 8-bit channels for red, green, and blue RGB and one 8-bit alpha channel. The alpha channel is really a *mask* -- it specifies how the pixel's colors should be merged with another pixel when the two are overlaid, one on top of the other.

B. Feathering:

A better approach to averaging is to weight pixels near the center of the image more heavily and to down-weight pixels near the edges. When an image has some cutout regions, down-weighting pixels near the edges of both cutouts and edges is preferable. This can be done by computing a *distance map* or *grassfire transform.* Weighted averaging with a distance map is often called *feathering* and does a reasonable job of blending over exposure differences however, blurring and ghosting can still be problems.

C. Pyramidal blending:

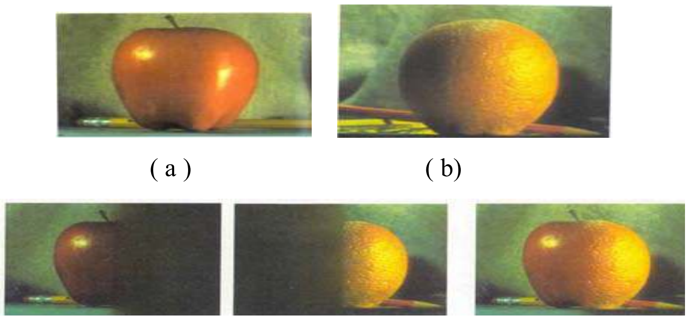
Once the seams have been placed and unwanted objects removed, we still need to blend the images to compensate for exposure differences and other mis-alignments. An attractive solution to this problem was developed by Burt and Adelson Instead of using a single transition width, a frequency adaptive width is used by creating a band-pass (Laplacian) pyramid and making the transition widths a function of the pyramid level. First, each warped image is converted into a band-pass (Laplacian) pyramid. Next, the masks associated with each source image are converted into a lowpass (Gaussian) pyramid and used to perform a per-level feathered blend of the band-pass images. Finally, the composite image is reconstructed by interpolating and summing all of the pyramid levels (band-pass images).

1) The Laplacian Pyramid:

The Gaussian pyramid is a set of low-pass filtered images. in order to obtain the band-pass images required for the multiresolution spline we subtract each level of the pyramid from the next lowest level. Because these arrays differ in sample density, it is necessary to interpolate new samples between those of a given array before it is subtracted from the next lowest array. Interpolation can be achieved by reversing the REDUCE process. We shall call this an E XPAND operation. Let $G_{l,k}$ be the image obtained by expanding *Gl, k* times. Then,

$$G_{l,0} = G_l,$$

An example shows the result of splining two quite different images, an apple and an orange (Figures 9a and 9b). The mosaic obtained without a spline is given in Figure 9c,9d, while that obtained with the spline is shown in Figure 14e.



(a) (b)

(c) (d) (e)

Fig9.Pyramidal Blending

V. Image Stiching Application:

Image Stiching has been proposed in the literature as a means for different applications. The main applications ce

1. texture synthesis,
2. object recognition
3. medical imaging.
4. To produce digital maps
5. To produce satellite photos

VI Conclusion:

Image stiching plays a vital role in 21st century which is emerging with various research programs & advances for more sophisticated & enjoyable life. Image enhancement has importance in entertainment. It evolves the enhancement of photographs, in medical field. image processing is used to produce digital maps & satellite photos , which yields more accurate results. This is a rapidly growing field & has a very wide scope.

REFERENCES

[1] P.J.Burt, E.H. Andelson, “A multiresolution spline with application to image mosaics”, ACM transactions on Graphics 2 (4) (1983).

[2] M. Brown and D. G. Lowe, “Recognizing Panorama”,-2003

[3] 3. Richard Szeliski, “Image Alignment and Stitching”-Microsoft Reasearch lab 2005.

[4] Jiaya jia, Chi-keung Tang, “Image Stitching using Structure Deformation”-IEEE transaction on pattern analysis and machine intelligence, vol 30-2008

[5] E. H. Adelson, C. H. Anderson, J. R. Bergen “Pyramid methods in image processing”-1984

[6] J.-Y. Guillemaut, J. Kilner, J. Starck, A. Hilton, “Dynamic Feathering: Minimising Blending Artefacts i

[7] n View-Dependent Rendering” - IEEE transaction on Image processing 2008..

[8] Jiaya Jia, Chi-Keung Tang, “Eliminating Structure and Intensity Misalignment in Image Stitching” Research Grant Council of Hong Kong Special Administration Region, China (AOE/E-1999).

[9] H.-Y. Shum and R. Szeliski. “Construction of panoramic mosaics with global and local alignment”. International Journal of Computer Vision, 36(2):101–130, February 2000. Erratum published July 2002,48(2):151-152.

[10] Xia Wan and C.-C. Jay Kuo” A New Approach to Image Retrieval With Hierarchical Color Clustering” ieee transactions on circuits and systems for video technology, september 1998