UNIT – V

Motion Models

The simplest possible motion model to use when aligning images is to simply translate and rotate them in 2D

- Used for Panoramic image stitching.
  - Done picture is assembled from several overlapping photographs. (ie) capture multiple images from different viewpoints
- In a panography, images are translated, rotated and scaled.
- also we can employ RANSAC or Hough Transform.

RANSAC → Random sample Consensus.

Image stitching algorithms create the high resolution photo-mosaics used to produce digital maps and satellite photos. They also come bundled with most digital cameras and can be used to create beautiful ultra wide angle panoramas.

Before we can register and align images, we need to establish the mathematical relationships that map pixel co-ordinates from one image to another. A variety of such

Parametric motion models are possible, from simple 2D transforms, to planar perspective models, 3D camera rotations, lens distor and mapping to non-planar (eg cylindrical) surfaces.

Types of model

1) Parametric motion describing the deformation of a planar surface as viewed from different positio can be described with an eight-parameter homograph
2) Camera undergoing a pure rotation induces a different kind of homography.
3. Spherical and cylindrical composing surfaces and show how, under favorable circumstances they can be used to perform alignment using pure translation.

Deciding which alignment model is appropriate for a given situation or set of data is a model selection problem.

The simplest motion model is Planar perspective motion model.

Application

* White board and document scanning

The simplest image-stitching application is to stitch together a number of image scans taken on a flatbed scanner.

Planar perspective motion

The simplest motion model to use when aligning images is to simply translate and rotate them in 2D. This is exactly the same kind of motion that we use if we had overlapping photographic prints. This is also the kind of technique to create the collages called as joiners.

Creating such collages, which show visible seams and inconsistencies is popular on web sites such as Flicker called as panography.

Translation and rotation are usually adequate motion models to compensate for small camera motions in applications such as photo and video stabilization and meiging.

consider the mapping between 2 cameras viewing a common plane can be described using a 3x3 homography.

Consider the matrix M_{10} that arises when mapping a pixel in one image to a 3D point and then back onto a second image

$$\tilde{x}_1 \backsim \bar{P}_1 \bar{P}_0^{-1} x_0 = M_{10} \tilde{x}_0$$

when the last row of the P_0 matrix is replaced with a plane equation $\hat{n}_0 \cdot p + c_0$ and points are assumed to lie on this plane. ie their disparity is $d_0 = 0$, we can ignore the last column of M_{10} and also its last row, since we do not care about the final Z-buffer depth.

The resulting homography matrix $\tilde{H}_{10}$ (the upper left 3×3 sub-matrix of M_{10}) describes the mapping between pixels in the 2 images

$$\tilde{x}_1 \backsim \tilde{H}_{10} \tilde{x}_0$$

Mean shift and Mode finding

Rotational Panoramas

The most typical case for panoramic image stitching is when the camera undergoes a pure rotation. All the points are very far from the camera ie on the plane infinity.

setting $t_0 = t_1 = 0$, we get the simplified 3x3 homography

$$\tilde{H}_{10} = K_1 R_1 R_0^{-1} K_0^{-1} = K_1 R_{10} K_0^{-1}$$

where $K_k = \text{diag}(f_k, f_k, 1)$ is the simplified camera intrinsic matrix assuming that $c_x = c_y = 0$. (ie) we are indexing the pixels starting from the optical center

This can also be re-written as

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \simeq \begin{bmatrix} f_1 & & \\ & f_1 & \\ & & 1 \end{bmatrix} R_{10} \begin{bmatrix} f_0^{-1} & & \\ & f_0^{-1} & \\ & & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix}$$

or

$$\begin{bmatrix} x_1 \\ y_1 \\ f_1 \end{bmatrix} \simeq R_{10} \begin{bmatrix} x_0 \\ y_0 \\ f_0 \end{bmatrix}$$

which reveals the simplicity of the mapping equations and makes all of the motion parameters explicit.

Thus, instead of the general 8-parameter homography relating a pair of images, we get the three-four- or five parameter 3D rotation motion models corresponding to the cases where the focal length f is known, fixed or variable.

Gap closing

* Used to estimate a series of rotation matrices and focal lengths, which can be chained together to create large panoramas. Because of accumulated errors there will invariably be either a gap or an overlap occurs in panography.

We can solve this pbm by matching the first image in the sequence with the last one. The difference blw the 2 rotation matrix estimates associated with the repeated first indicates the amount of misregistration.

This error can be distributed evenly across the whole sequence by taking the quotient of the 2 quaternions (alternate way to describe orientation or rotation in 3D space using an ordered set of 4 numbers) associated with these rotations and dividing this "error quaternion" by the no. of images in the sequence.

we can also update the estimated focal length based on the amount of misregistration.

To do this, we first convert the ~~estimated focal length based on the~~ error quaternion into a gap angle θ_g and then update the focal length using the equation

$$f' = f(1 - \theta_g/360^\circ)$$

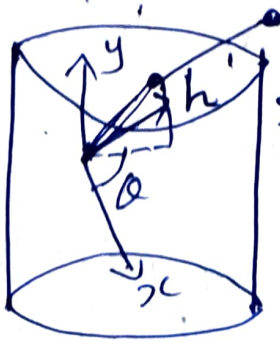
cylindrical and spherical co-ordinates.

* An alternative to using homographies or 3D motions to align images is to first warp the images into cylindrical co-ordinates and then use a pure translational model to align them.

* Assume that the camera is in its canonical position (ie) its rotation matrix is the Identity. $R = I$. so that the optical axis is aligned with the Z axis and the y axis is aligned Vertically.

So 3D ray corresponding to an (x,y) pixel is (x,y,f).

We wish to project the image into cylindrical surface of unit radius.



Points on this surface are parameterized by an angle θ and a height h, with the 3D cylindrical co-ordinates corresponding to (θ, h) is given by

$(\sin\theta, h, \cos\theta) \propto (x,y,f)$

From this correspondence we can compute the formula for the warped or mapped (bent) co-ordinates.

$$x' = s\theta = s\tan^{-1}\frac{x}{f} \qquad f \rightarrow \text{focal length}$$

$$y' = sh = s\frac{y}{\sqrt{x^2+f^2}}$$

Where s is an arbitrary scaling factor (sometimes called the radius of the cylinder) that can be set to s=f to minimize to distortion near the center of the image.

* Images can also be projected onto a spherical surface, which is useful if the final panorama includes a full sphere or hemisphere of views instead of a cylindrical strip.

P = (x, y, z).

The sphere is parameterized by 2 angles (θ, φ)

(sinθcosφ, sinφ, cosθcosφ) ∠ (x, y, f)

$$x' = s\theta = s\tan^{-1}\frac{x}{f}$$

$$y' = s\phi = s\tan^{-1}\frac{y}{\sqrt{x^2+y^2}}$$

Another Co-ordinate mapping is Polar mapping. where the north Pole lies along the Optical axis rather than the Vertical axis.

$(\cos\theta\sin\phi, \sin\theta\sin\phi, \cos\phi) = s(x, y, z)$

Mapping equations.

$$x' = s\phi\cos\theta = s\frac{x}{r}\tan^{-1}\frac{r}{z}$$

$$y' = s\phi\sin\theta = s\frac{y}{r}\tan^{-1}\frac{r}{z}.$$

where $r = \sqrt{x^2+y^2}$ is the radial distance in the (x, y) Plane

Bundle adjustment -

One way to register a large no. of images is to add new images to the panorama one at a time, aligning the most recent image with the previous ones already in the collection and discovering which images it overlaps. In the case of 360° panoramas, accumulated error may lead to the

Presence of a gap (or excessive overla
b/w the 2 ends of the panorama, which c
be fixed by stretching the alignment
all the images using a process called gap closing

An alternative is to simultaneously align all the images using a least squares framework to correctly distribute any mis-registration errors.

The process of simultaneously adjusting pose parameters for a large collection of overlapping images is called bundle adjustment. (global alignment)

we use the concept feature based approach.

Feature based alignment equation

$$E_{pairwise\_LS} = \sum_i \|r_i\|^2 = \|\tilde{x}_i'(x_i;p) - \hat{x}_i'\|^2$$

↓ left at the end of a Proces

$r_i = \hat{x}_i' - \tilde{x}_i'$ is the residual b/w the measured location $\hat{x}_i'$ and its corresponding current predicted location $\tilde{x}_i'$

Parallax removal

Once we have optimized the global orientations and focal lengths of our cameras, we may find that the images are not perfectly aligned: ie resulting stitched image looks blurry on some places.

This can be caused by a variety of factors, including unmodeled radial distortion, 3D parallax (failure to rotate the camera around its optical center), small scene motions such as waving tree branches, and large scale scene motions such as people moving in and out of pictures.

3D parallax can be handled by doing a full 3D bundle adjustment (ie)

Parallax

Parallax is difference in the apparent position of ~~an~~ object viewed along 2 different lines of sight.

when the motion in the scene is very large (ie) when objects appear and disappear completely, a soln is to simply select pixels from only one image at a time as the source for the final composite.

For multi-image alignment, instead of having a single collection of pairwise feature correspondences $\{x_i, \hat{x}'_i)\}$, we have a collection of n features, with x_{ij} denoting the location of the ith feature point in the jth image and its scalar confidence denoted by c_{ij}.

Each image also has some associated pose parameters.

We assume that this pose consists of a rotation matrix R_j and a focal length f_j.

The equation mapping a 3D point x_i into a point x_{ij} in frame j can be re-written as

$$\tilde{x}_{ij} \sim K_j R_j x_i \quad \text{and} \quad x_i \sim R_j^{-1} K_j^{-1} \tilde{x}_{ij},$$

where $K_j = \text{diag}(f_j, f_j, 1)$ → calibration matrix.

The motion mapping a point x_{ij} from frame j into a point x_{ik} in frame k is

$$\tilde{x}_{ik} \sim H_{kj} x_{ij} = K_k R_k R_j^{-1} K_j^{-1} \tilde{x}_{ij}.$$

When the motion is small (on the order of a few pixels), general 2D motion estimation (optical flow) can be used to perform an appropriate correction before blending using a process called

Local alignment.

Local alignment starts with the global bundle alignment used to optimize the camera poses. Once these have been estimated, the desired location of a 3D point x_i can be estimated as the average of the back projected 3D locations.

$$\bar{x}_i = \sum_j c_{ij} \tilde{x}_d(\hat{x}_{ij}; R_j, t_j) \Big/ \sum_j c_{ij}$$

which can be projected into each image j to obtain a target location $\bar{x}_{ij}$.

The difference between the target locations $\bar{x}_{ij}$ and the original features x_{ij} provide a set of local motion estimates

$$u_{ij} = \bar{x}_{ij} - x_{ij}$$

Another approach to motion-based deghosting is, estimate dense optical flow b/w each i/p image and a central reference image.

The accuracy of the flow vector is checked using a photo-consistency measure before a given warped pixel is considered valid and is used to compute a high dynamic range radiance estimate, which is the goal of the algorithm.

Photo consistency.

It is a scalar function that measures the visual compatibility of a 3D reconstruction with a set of calibrated images.

Optical flow

It is the motion of objects b/w the consecutive frames of the sequence, caused by the relative motion b/w the camera and the object.

Dense optical flow

Dense optical flow Computes the optical flow vector for every pixel of the frame which may be responsible for its slow speed.

Compares 2 images to estimate the apparent motion of each pixel in the one of the images.

Recognizing Panoramas.

To Perform fully automated image stitching is a technique to recognize which images actually go together. called 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.

However, users often jump around when taking panoramas, e.g they may start a new row on top of a previous one, jump back to take a repeat 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.

To recognize panoramas, first find all pairwise image overlaps using a feature-based method and then find connected components in the overlap graph to "recognize" individual panoramas.

The feature based matching stage first extracts scale invariant feature transform (SIFT) feature locations and feature descriptors from all the input images and places them in an indexing structure.

For each image pair under consideration, the nearest matching neighbour 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 used to find a set of inlier matches; pairs of matches are used to hypothesize similarity motion model that are then used to count the number of inliers.

Compositing

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 ghoting.

Choosing a compositing surface

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 its reference co-ordinate system. The resulting composite is sometimes called a flat panorama, since the projection onto the final surface is still a perspective projection, and hence straight lines remain straight.

For larger fields of view, we cannot maintain a flat representation without excessively stretching pixels near the border of the image.

The usual choice for compositing larger panoramas is to use cylindrical

Or Spherical projection.

In fact, any surface used for environment mapping in computer graphics can be used, including a cube map, which represents the full viewing sphere with the six square faces of a cube.

Recent development in Panoramic photography has been the use of stereographic projections looking down at the ground (in an outdoor scene) to create "little planet" renderings.