Advanced Computer Vision
Final for Spring 2022

Course#: 922 U3910::CSIE7421 Date: May 31, 2022

Your ID#: 61047086s Instructor: Chiou-Shann Fuh

Your Name: Rohit Das

1. (48%) Please define the following terms and explain the content, purpose, and application of each term and give an illustrative example if possible. If possible, define the term in mathematical equation. For example:
thresholding: an image point operation that produces a binary image from a gray scale image. A binary-1 is produced on the output image whenever a pixel value on the input image is above a specified minimum threshold level. A binary-0 is produced otherwise. Alternatively, thresholding can produce a binary-1 on the output image whenever a pixel value on the input image is below a specified maximum threshold level. A binary-0 is produced otherwise.

(1) image alignment:

Feature-based alignment of multiple images by estimating the motion between two or more sets of matched 2D or 3D points.

(2) image stitching:

Image stitching algorithms create the high-resolution photo-mosaics used to produce today's digital maps and satellite photos. They are also now a standard mode in smartphone cameras and can be used to create beautiful ultra-wide-angle panoramas. Image stitching originated in the photogrammetry community, where more manually intensive methods based on surveyed ground control points or manually registered tie points have long been used to register aerial photos into large-scale photo-mosaics (Slama 1980). While early techniques worked by directly minimizing pixel-to-pixel dissimilarities, today's algorithms extract a sparse set of features and match them to each other.

(3) spherical panorama:

To generate cylindrical or spherical panoramas, we need to take images from a rotating camera in such way that it can take a series of upto 50% overlapped images and then:

1.Estimate the amount of radial distortion by taking some pictures with lots of long straight lines near the edges of the image and then using the plumb-line method

2. Compute the focal length either by using a ruler and paper or by rotating your camera on the tripod, overlapping the images by exactly 0% and counting the number of images it takes to make a 360° panorama.
3. Convert each of your images to cylindrical coordinates.
4. Line up the images with a translational motion model using either a direct pixel-based technique, such as coarse-to-fine incremental or an FFT, or a feature-based technique.
5. (Optional) If doing a complete 360° panorama, align the first and last images. Compute the amount of accumulated vertical misregistration and re-distribute this among the images.
6. Blend the resulting images using feathering or some other technique.

(4) Jacobian:

A matrix of partial derivatives. To obtain sub-pixel estimates, the more commonly used approach is to perform gradient descent on the SSD (Sum of Squared Difference) energy function using a Taylor expansion of the image function. Which on further calculation will result in Jacobian or image gradient function.

(5) bundle adjustment:

The most accurate way to recover structure and motion is to perform robust non-linear minimization of the measurement (re-projection) errors, which is commonly known in the photogrammetry (and now computer vision) communities as bundle adjustment

(6) inverse compositional alignment (Baker-Matthews):

The inverse compositional approach supports group wise geo- metric transformations, and it improves efficiency by performing most computationally expensive calculations (i.e. the Gauss-Newton approximation to the Hessian matrix) at the pre-computation phase.

(7) template matching:

Template matching is a technique in digital image processing for finding small parts of an image which match a template image. It can be used in manufacturing as a part of quality control, a way to navigate a mobile robot, or as a way to detect edges in images.

(8) computational photography:

The process of creating new images from one or more input photographs, often

based on the careful modeling and calibration of the image formation process is known as Computational Photography. Computational photography techniques include merging multiple exposures to create high dynamic range images, increasing image resolution through blur removal and super-resolution and image editing and compositing operations.

(9) structure from motion:

One of the neatest applications of structure from motion is to estimate the 3D motion of a video or film camera, along with the geometry of a 3D scene, in order to superimpose 3D graphics or Computer-Generated Images (CGI) on the scene.

For very small motions, or motions involving pure camera rotations, one or two tracked points can suffice to compute the necessary visual motion. For planar surfaces moving in 3D, four points are needed to compute the homography, which can then be used to insert planar overlays.

(10) SLAM:

SLAM means that when a certain mobile device (such as a robot) starts from an unknown location in an unknown environment, it uses sensors to observe and locate its own position, posture, and trajectory during the movement and then builds an incremental map based on its own position, to achieve the purpose of simultaneous positioning and map construction.

(11) geometric intrinsic calibration:

The computation of the internal (intrinsic) camera calibration parameters can occur simultaneously with the estimation of the (extrinsic) pose of the camera with respect to a known calibration target. This, indeed, is the "classic" approach to camera calibration used in both the photogrammetry (Slama 1980) and the computer vision (Tsai 1987) communities.

(12) pose estimation:

A particular instance of feature-based alignment, which occurs very often, is estimating an object's 3D pose from a set of 2D point projections. This pose estimation problem is also known as extrinsic calibration, as opposed to the intrinsic calibration of internal camera parameters such as focal length. The problem of recovering pose from three correspondences, which is the minimal amount of information necessary, is known as the perspective-3-point-problem (P3P),2 with extensions to larger numbers of points collectively known as PnP

(13) depth estimation:

Stereo Matching / Depth Estimation is the process of taking two or more images and building a 3D model of the scene by finding matching pixels in the images and converting their 2D positions into 3D depths. Depth estimation algorithms can convert a pair of color images into a depth map.

(14) epipolar line:

The line connecting two camera points is known as epipolar line.

(15) rectification:

Reproject image planes onto a common plane parallel to the line between camera centres

(16) sparse correspondence:

Feature-based matching algorithms searched for corresponding location in the other images by using either interest operators or edge detectors. There is also a desire to match scenes with different illumination in some application.

(17) Lambertian surface:

The diffuse component (also known as Lambertian or matte reflection) scatters light uniformly in all directions and is the phenomenon we most normally associate with shading, e.g., the smooth (non-shiny) variation of intensity with surface normal that is seen when observing a statue

(18) level of detail:

Once a triangle mesh has been created from 3D data, it is often desirable to create a hierarchy of mesh models, for example, to control the displayed level of detail (LOD) in a computer graphics application. (In essence, this is a 3D analog to image pyramids. One approach to doing this is to approximate a given mesh with one that has subdivision connectivity, over which a set of triangular wavelet coefficients can then be computed.

(19) shape from focus:

A strong cue for object depth is the amount of blur, which increases as the object's surface moves away from the camera's focusing distance. Moving the object surface away from the focus plane increases the circle of confusion, according to a formula that is easy to establish using similar triangles

(20) shape from texture:

The variation in foreshortening observed in regular textures can also provide useful information about local surface orientation. Shape from texture algorithms require a number of processing steps, including the extraction of repeated patterns or the measurement of local frequencies to compute local affine deformations, and a subsequent stage to infer local surface orientation.

(21) view interpolation:

(22) layered depth image:

An LDI is a view of the scene from a single input camera view, but with multiple pixels along each line of sight. The size of the representation grows only linearly with the observed depth complexity in the scene. Moreover, because the LDI data are represented in a single image coordinate system, McMillan's warp ordering algorithm can be successfully adapted. As a result, pixels are drawn in the output image in back-to-front order. No z-buffer is required, so alpha-compositing can be done efficiently without depth sorting. This makes splatting an efficient solution to the resampling problem.

(23) photo tourism:

Photo tourism is a system for browsing large collections of photographs in 3D. Our approach takes as input large collections of images from either personal photo collections or Internet photo sharing sites, and automatically computes each photo's viewpoint and a sparse 3D model of the scene. Our photo explorer interface enables the viewer to interactively move about the 3D space by seamlessly transitioning between photographs, based on user control.

(24) sprites:

An image sprite is a collection of images put into a single image.

2. (6%) Please describe the method, steps, and results of Magnet Defect Inspection.

Method - HSV Black/ White threshold/ closing

Step - Take a photo of the sample -> image processing finds defect contour -> present results

Result - Find magnets defects

3. (6%) Please describe the method, steps, and results of Lens Flare Detection for Camera Module.

Method – Detect Flare

Steps - largest channel in each pixel minus the average value of the three channels as the value of the grayscale image -> Constantly take the threshold value and divide the image into several layers using the brightness of different intervals - > find the edge contours in the image.

Result – Flare Detected in output image

4. (4%) Please explain in detail two key ideas of Horn-Schunck optical flow.

- Brightness Constancy
- Small motion
- Smooth Flow
- Global method (Dense)

5. (6%) Please describe the method, steps, and results of Rethinking BiseNet for Real-Time Semantic Segmentation.

Method: Rethinking BiSeNet For Real-time Semantic Segmentation.

Step:

1. Use STDC module to replace the context path backbone in BiSeNet.
2. Use a Detail Aggregation module by integrating spatial information in single-stream manner.

Result: Faster, more precise.

6. (6%) Please list 6 steps and explain in detail the process of generating High Dynamic Range images from the multiple exposure photographs.

- Align photographs when necessary.
- Estimate the radiometric response function.
- Estimate radiance map.
- Perform motion compensation (anti-ghosting) when necessary.
- Perform Tone Mapping to reduce bit depth.
- Export final image in the commercial format, like JPG.

7. (6%) Please describe the method, steps, and results of Noise Reduction for Surgical Video Recording System.

Methods:

- Noise characterisation: measure spatial variance over the regions
- Noise reduction:
  - Median
  - Bilateral
  - Nonlocal Means Denoising
- Steps:
  - Acquire images of the calibration target
  - Measure noise and spatial resolution
  - Apply noise reduction filters
  - Measure resulting noise and spatial resolution
  - Collect and record data

Result: noise reduction for the surgical imaging system.

8. (6%) Please describe the method, steps, and results of Driver Monitoring System with Embedded System KL520.

Method – Driver monitoring system using KL 520

Steps – Camera input -> Driver monitoring system (Object Detection)-> Detect Behavior -> Alert the Driver

Output -> Alert the Driver depending on the unfavorable behavior

9. (6%) Please describe the method, steps, and results of 3D Solder Ball Reconstruction with Simultaneous Algebraic Reconstruction Technique and X-Ray Images.

Method - Filtered back projection on Sinogram image

Steps – Input Image > Create Filter > Fourier transform > Filter > Inverse Fourier Transform > Back projection > Normalization > Output

Output - Reconstructed image from sinogram

10. (6%) Please describe the method, steps, and results of Parameter Recommendation for Image Quality of Smartphone Camera.

- Purpose: Reduce tuning time
- Methods:
  - Phase 1: model the relation between ISP parameter x and IQM score y
  - Phase 2: find optimal ISP parameter corresponding to the desired IQM score
- Result: Recommend ISP parameter successfully

Bonus

11.(2%) Please write the Chinese name of Professor Chiou-Shann Fuh.

傅楸善教授

12.(2%) What's Professor Chiou-Shann Fuh's pet phrase?

(a)酷斃了 Cool (b)帥呆了 Handsome (c)好極了 Good (d)棒透了 Awesome

Good

13.(2%) Is text or joke more important?

Joke