Features

Colour Features: It is a property of a single pixel and preserve colour constancy. Colour histograms are invariant to translation and rotation. Specifically, colour histograms summarize target objects.

Texture Features: capture the frequency with which patterns of colour/grey level appear. Local Binary Patterns: 1. Divide the patch into cells 2. Compute the local patch description number of each pixel. 3. Histogram these numbers 4. Optionally normalize each histogram 5. Concatenate histograms to make the feature vector.

Shape Features: Focus on image gradient measures: Distributions of gradients and gradient orientations reflect boundary shape. Histogram of Oriented Gradients(Hog): 1.Divide the patch into cells 2. Define larger blocks, covering several cells. 3. Compute gradient magnitude and orientation at each pixel. 4. Compute a local weighted histogram of gradient orientations for each cell. 5. Concatenate histogram to form a HoG vector for each block 6. Normalize vector values by dividing by some function of vector length

SIFT (Scale Invariant Feature Transform)

Scale Invariance: Find points whose surrounding patches are distinctive with using Gaussian mask

Translation Invariance: key point localization. (Scale alone gives too many points, some not accurate) if shifting the window causes a big change, its more likely to be a good feature.

Rotation Invariance: To handle rotation, find the dominant orientation of the image patch and rotate the patch according to the angle, which is given by x+. the eigenvector of H corresponding to the larger eigenvalue.

Image Stitching

Combine two or more overlapping images to make one larger image. Need to geometric transformation on image to fit on another image. Similarity transform ( 4 Degree of Freedom) = transformation + rotation + scale. Affine transform ( 6 DoF) = translation + rotation + scale + aspect ratio + shear Homography = 8 DoF

The order: 1. Take a sequence of images from the same position. 2. compute transformation on two images (extract feature points(SIFT), find good matches (compare feature vectors)) 3. Shift the second image to overlap with the first 4. Blend together 5. Repeat 2-4

Least Squares Formulation : minimize sum of squared residuals

RANSAC: Random sample consensus 1. Select four feature pairs(at random) 2. Compute homography H(exact) 3. Compute inliers ( keep the largest set of inliers, Re-compute least-squares H estimate using all of the inliers)

Warping(shifting) : Once we found the transformation, we shift one image according to the transformation.

Blending: feathering with ramp, alpha blending, pyramid blending, and multiband blending

Panoramas: 1. Extract SIFT points, descriptors from all images 2. Find KNN for each point 3. For each image, select M candidate matching images by counting matched keypoints, solve homography Hij for each matched image, finally decide if match is valid. 4. Make a graph of matched pairs and find connection 5. Solve for rotation and camera parameter and render with multiband blending

Optical Flow

The Motion Field vs. Optical Flow: The motion field: assign a velocity vector to each point in the image. Optical Flow: apparent motion of the brightness pattern in an image.

Aperture problem: refers to the fact that the motion of a one-dimensional spatial structure, such as a edge, cannot be determined unambiguously if it is viewed through a small aperture such that the ends of the stimulus are not visible. With aperture problem, we know it has to move to a certain point, but do not know on which direction.

Horn-Schunk Method vs. Lucas-Kanade: Both assume brightness constancy and smooth flow.

Horn-Schunk Method: Global information / Dense, smooth flow field BUT errors can propagate / iterative and slow / object boundaries not sharp

Lucas-Kanade: local process / Easy and fast calculation / more stable in noise BUT Errors on boundaries.

Both has same constraint: the flow field is smooth and neighbor pixels in the image should have similar optical flow.

Image Pyramids: A series of images. Each is smaller than the last. Each pixel is a smaller image corresponds to a number of pixels in the larger one. Images are often blurred before being reduced. Up are lower resolution but can be processed faster.

Pyramids allow us to use optical flow for large motion.

Dense vs. Sparse optical flow: Dense: compute estimate for each pixel, higher accuracy at the cost of slow/expense computation. Sparse: Compute estimate for some good feature points(SIFT), mush less computation cost.

Camera

Pinhole Camera: Add a barrier to block off most of the rays to reduce blurring(aperture).

Lenses: A lens focuses parallel rays onto a single focal point.

Perspective projection vs. Orthographic projection: In the perspective view, objects which are far away are smaller than those nearby. In the orthographic view, all objects appear at the same scale. Perspective viewpoints give more information about depth and are often easier to view because we use perspective views in real life.

Extrinsic vs. Intrinsic parameters: The extrinsic represent the location of the camera in the 3D scene. The intrinsic represent the optical center and focal length of the camera. The world points are transformed to camera coordinates using the extrinsic parameters.

Stereo

Depth from stereo: Recover depth by finding image coordinate x’ that corresponds to x.

Epipolar constraint: x’ always lies on epipolar line in the other image. Epipolar lines are intersections of epipolar plane with image planes (always come in corresponding pairs) Epipolar line can be computed in both calibrated and uncalibrated case.

Rectification: Reproject image planes onto a common plane parallel to the line between the line between camera center. Pixel motion is horizontal after this transformation. Two homographies, one for each input image reprojection.

Correspondence search

3D Reconstruction

Structured light: Project any spatio-temporal pattern of light on the object surface. Avoid problems of 3D estimation in scenes with complex textures. Make finding correspondence very easy because it contains multiple layers of light compared to stereo.

Photometric stereo: Estimate the surface normals of a given scene given multiple 2D images taken from the same viewpoint, but under different lighting conditions. Basic photometric stereo required a Lambertian reflectance model: pixel intensity = diffuse albedo constant(=reflection coefficient)(normal) \* lighting direction

3D from photographs: practical, fast, non-intrusive, low cost, easily deployable outdoor, but low accuracy and results depend on material properties

Space Carving Algorithm: 1. Initialize to a volume V containing the true scene 2. Choose a voxel on the outside of the volume 3. Project to visible input images 4. Carve if not photo-**consistent** 4. Repeat until convergence

Bag of Features

Bag of features methods analyze the large set of very specific features generated by a training set of images and identify a small set of useful, more generic features.

K-mean Clustering: Minimize distances between points and their nearest cluster centers: allows user to choose vocabulary size. If vocab is too small, visual words not representative of all patches. If too large, quantization artifacts, overfitting

Histograms of these abstract features provide compact representations of images

Histograms can be weighted by inverse document frequency, the feature vectors supplied to a classifier, and matched to extract images of a given object from a large-scale database.

Object Recognition

3 important recognition problems: recognition: identify the main object in an image, Detection: also find the object’s image location, Pose: find the object’s 3D, real world location.

Early works: pictorial structure, Classical Dalal & Triggs

Noadays: Statistical learning framework, feature representation, and trainable classifiers

Viola Jones Method(Recognition)

Developed for face recognition, but general. Slide a window across image and evaluate a face model at every location. Key ideas: Integral images for fast feature evaluation, boosting for feature selection, Attentional cascade for fast rejection of non-face windows. Integral image makes feature extraction faster and allows consideration of more features. The integral image computes a value at each pixel(x,y) that is the sum of the pixel values above and to the left of (x,y). This can quickly be computed in one pass through the image.

Three key contributions: Integral images( basis of SURF), Boosting, and Cascaded classifiers

Together they allow automatic selection and learned use of features from a previously unfeasibly large set.

CNN Convolutional Neural Networks

Neural networkds that use convolution in place of general matrix multiplication in at least one of their layers.

ResNet: The residual module, introduce skip or shortcut connections( existing before in various forms in literature), make it easy for network layers to represent the identity mapping.

Convolutional Neural networks take feature learning much further, providing an entirely learned solution

GAN Generative Adversarial Networks

Two neural networks contest with each other in game( in the form of a zero-sum game, where one agent’s gain is another agent’s loss) Given a training set, this technique learns to generate new data with the same statistics as the training set. For example, a GAN trained on photographs can generate new photographs that look at least superficially authentic to human observers, having many realistic characteristics