

CS585 Assignment 5 Written Part

Yida Xin

December 10, 2018

Exercise 1: Tracking – Data Association and Kalman Filters

- (a) Because Algorithm 1 gives a higher value than Algorithm 2, my judgment is that Algorithm 1 is the Greedy Method, because the Greedy Method solely greedily minimizes the sum of distances, and that Algorithm 2 is the Hungarian Method, because it seems to have taken directional information into account and produced the actual optimal sum.
- (b) Yes. Because Kalman Filter takes directional information into account, both Greedy Method and Hungarian Method could produce 8.8 if they both use Kalman Filter for their filtering step. However, if they both use the more naïve Alpha-Beta Filter, then they may not both produce 8.8.
- (c) See blue pen in original homework sheet.
- (d) See black pen in original homework sheet.
- (e) Kalman Filter minimizes the **a posteriori expected value of error covariance**, i.e. the difference between the predicted and the detected values.
- (f) MHT builds a hypothesis tree. Consequently, “pruning” could mean both pruning the width/span of the tree and pruning the depth of the tree.

Hypothesis-oriented pruning (a.k.a. width pruning): only keep a fixed number of the “best” hypotheses formed based on the current measurements.

Track-oriented pruning (a.k.a. depth pruning): whenever a track fails to continue to generate more than one hypothesis at whichever time step, delete that entire track starting from its beginning.

Exercise 2: Optical Flow

For the entire Exercise 2, see attached drawing.

Exercise 3: 3D Rotation and Absolute Orientation

- (a) Three parameters—roll (x-axis), yaw (y-axis), pitch (z-axis)—are needed in order to describe an arbitrary rotation in the 3D space.

(b) The following matrix

$$\begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

represents the θ -parameterized rotation in the x - y plane in the 3D space.

(c) In order to represent a 3D rotation, a quaternion must be of *unit* length and we must “find a way of mapping purely imaginary quaternions into purely imaginary quaternions in such a way that dot products are preserved, as is the sense of dot products.

(d) The angle between $(1, 0, 1)$ and $(2, -1, 0)$ is

$$\arccos \left(\frac{(1, 0, 1) \cdot (2, -1, 0)}{\|(1, 0, 1)\| \|(2, -1, 0)\|} \right) = \arccos \left(\frac{2}{\sqrt{10}} \right)$$

According to what I have found on the Wikipedia page “Quaternions and Spatial Rotations” under the section “Quaternion-derived rotation matrix,” the 3×3 rotation matrix that corresponds to the quaternion $(0, 1, 0, 0)$ is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix},$$

which represents the 180-degree rotation around the x -axis. Thus, the vectors $(1, 0, 1)$ and $(2, -1, 0)$ simply become

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

The angle between $(1, 0, -1)$ and $(2, 1, 0)$, then, is

$$\arccos \left(\frac{(1, 0, -1) \cdot (2, 1, 0)}{\|(1, 0, -1)\| \|(2, 1, 0)\|} \right) = \arccos \left(\frac{2}{\sqrt{10}} \right),$$

as desired.

(e) It is globally optimal, as the problem is formulated as a convex-optimization problem where a global optimal solution exists, and the closed form of that solution has been found.

(f) An example could be Rigid-Body Alignment/Registration/Orientation applied to medical imaging, e.g. aligning a patient’s lung image with a healthy lung image so as to identify the location of potential tumor(s). In this case, the input would be the so-called “landmarks,” i.e. selected points in both the model image and the test image that are supposed to make identification easy; the output would be the rotation matrix \mathbf{R} and the transition vector $\vec{r}_0 \in \mathbb{R}^3$.

Exercise 4: Stereoscopy

According to the setup, we have $b = 200$, $f = 50$, and pixel width/height = 0.013, all in millimeters.

- (a) The disparity, then, is $\delta = 12 \times 0.013 = 0.156\text{mm}$. That means $Z = bf/\delta = 200 \cdot 50/0.156 \approx 64102\text{mm} = 64.102\text{m}$.
- (b) There are no epipolar lines, because there are no epipoles, which in turn is because the two cameras share the same image plane.
- (c) There are no epipoles, because the line that connects the CoPs do not intersect the image plane, which in turn is because the optical axes are parallel.
- (d) Now that there is an angle between the two optical axes, there is also an angle between the two image planes, so the epipoles will be the intersections of the line that connects the CoPs and the two image planes, and the epipolar line for either camera is the distance between the intersection of the perpendicular line from the CoP to the image plane and the epipole.

Exercise 5: Object Recognition, Deep Learning, Face Recognition

- (a) The four image attributes that were varied in ILSVRC are (1) amount of texture, (2) color distinctiveness, (3) shape distinctiveness, and (4) real-world size. More details can be found in [<http://image-net.org/download-attributes>].
- (b) The leaders in this competition are SuperVision (SV) and Oxford_VGG (VGG). More “modern” variations of both architectures have come about since 2012, and these variations are still very much actively used in Computer Vision today.
- (c) Terminology-wise, one-to-one is called Face Verification while one-to-many is called Face Identification.
- (d) Common data-preprocessing techniques used in Deep Learning-based Face Recognition systems are dataset balancing, data normalization (a.k.a. feature rescaling), and data augmentation (a.k.a. dataset enlarging). All these preprocessing techniques aim to eliminate biases in the given dataset(s) — after all, neural net is a curve-fitter, so the more unbiased the input data are, the better.. An example where Data Balancing is needed is a dataset with which we want to train a neural net to reliably classify “dog” and “cat” — usually, a dataset doesn’t have, say, exactly 50% dog images and 50% cat images, alongside things like different poses and background noise. Data normalization is an attempt to get all the data points “on the same scale,” e.g. zero-mean normalization. Data augmentation is a method to artificially create more data points when there are not enough to begin with, using tricks like translation, rotation, flipping, etc. Neural nets are data hungry, so the more data the merrier.
- (e) A one-hot vector is a vector that has “1” in the desired entry, and “0”s in all the other entries. If we have a bunch of items of interest that we want to encode using vectors of the same dimensionality, then we index these items, initialize zero-vectors for them whose dimensions are the same as the total number of items, and then for each item we replace the “0” in the entry indexed by that item’s index with “1”.

- (f) In the context of Face Recognition, distractor images are images taken of people who are not of interest — images of “strangers.”
- (g) Supervised, sadly.
- (h) Deformable Parts Model (DPM).
- (i) “Bag” means *multiset*. A multiset is like a regular set, but can contain multiple occurrences of the same element. This is important in representing visual features. Just like in texts where the same word can repeat over and over again, the same (or roughly the same) visual features might occur in an image over and over again. A case in point, there would be multiple pixels in one image that have exactly the same RGB-value.
- (j) For train & test sets, usually we split the available dataset 70/30, usually randomly; we train the neural net on the 70, and then find out how well it has “learned” from those 70 by testing it on the remaining 30.

For train, validation, and test sets, usually we split 70/15/15 or something like that, usually randomly; we train the neural net on the 70, then run it on the middle 15 to find out both how good the learned parameters and how good the manually initialized hyperparameters are, with a focus on the latter, and then we adjust those hyperparameters, if necessary, and test the model on the test set.

Exercise 6: Lens Systems

- (a) True. This is a fact from physics.
- (b) False, only because you didn’t say “ideal” lens. For an ideal lens, nonetheless, this is true.
- (c) True. For a fixed magnification, decreasing the aperture diameter increases the f-number, which increases the depth of field.

Exercise 7: Photometric Stereo

- (a) True. This is the definition of Lambertian Surface.
- (b) This question.....I take it that I can fill in the rest with the words “vectors lengths.” If that’s the case, then this is true, because this is what the brightness-measurement formula dictates — that both the normal vector and the light-source vector be normalized.
- (c) True. A dot product of 1 (the value of E_2) stipulates the the surface normal vector \vec{n} must overlap \vec{s}_2 , and \vec{s}_2 is already given.
- (d) False. There exist an entire unit-radius disk whose angle with \vec{s}_1 is $\arccos(0.707)$ and whose angle with \vec{s}_2 is $\arccos(0.57)$. We need more information to determine which specific unit vector on that unit disk we are in fact looking for.