CPSC 425   **Stereo, Motion and Optical Flow**   **18 (Term 2)**
Practice Questions
Solutions

Multiple Part True/False Questions. For each question, indicate which of the statements, (A)–(D), are **true** and which are **false**? Note: Questions may have zero, one or multiple statements that are true.

Question 1. Consider conditions under which an epipolar constraint used in stereo matching holds between images from two cameras. Which of the following conditions are **true**? Which are **false**? Note: You can assume that the cameras perform standard perspective projection.

(A) The two cameras must have coplanar projection planes.

(B) The two cameras must face in the same direction (i.e., have parallel optical axes).

(C) The two images must be rectified.

(D) There are no restrictions on camera locations or orientations, an epipolar constraint always applies.

Solution : (A) False, (B) False, (C) False, (D) True.

Question 2. Stereo matching can be performed by correlating windows of pixels between the two images. But, it is difficult to know what window size to use. The following statements identify problems when the selected window size is too large. Which are **true**? Which are **false**?

(A) There will be more false matches due to ambiguity and image noise.

(B) The exact location of correct matches will be known with less accuracy.

(C) Places where depth is discontinuous will be poorly matched.

(D) The epipolar constraint is not as effective to limit the number of matches.

Solution : (A) False, (B) True, (C) True, (D) False.

Question 3. The Lucas–Kanade method makes several assumptions about motion and optical flow. Which of the assumptions, (A)–(D), are **true** of Lucas–Kanade and which are **false**? Note: This is a question about the Lucas–Kanade method, not about assumptions that may or may not be true, in general, about the world.

(A) Corresponding points in a sequence of images of a moving object have exactly the same brightness values.

(B) Sampling in x, y and t is frequent enough that the partial derivative, I_x, I_y and I_t, are well-defined

(C) The motion, $[u, v]$, is constant in the selected window about each image point, $[x, y]$.

(D) The matrix

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} \sum I_x{}^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y{}^2 \end{bmatrix}$$

has rank 2 in the selected window about each image point.

Solution : (A) True, (B) True, (C) True, (D) True

Short Answer Questions.

Question 4. The second edition of Pat Winston's textbook, *Artificial Intelligence*, published by Addison-Wesley, contains a discussion of stereo vision. Included is an extended example based on a stereo pair of images shown in the text as a figure. The figure caption reads, in part, "The two pictures are arranged so that you can see depth yourself with the aid of a stereoscopic viewer." At the last minute, prior to printing, a graphic designer at Addison-Wesley made the artistic decision that the stereo pair looked better arranged above and below (i.e., top to bottom) rather than left to right. Accordingly, that is how the initial press run was printed – a left/right stereo pair printed with the left image above and the right image below.

Winston was not amused and insisted that Addison-Wesley reprint the entire book again, at its cost, with the figure in question corrected. Aside: This is a true story.

Briefly describe why Winston would insist that the figure be corrected.

Solution : For (human) stereoscopic viewing, the epipolar lines need to be collinear. In a standard left/right stereo pair the epipolar lines are the corresponding (horizontal) scan lines. Thus a left/right stereo pair needs to be viewed with the left image on the left and the right image on the right, with corresponding scan lines collinear. (Aside: With the left image on the right and the right image on the left, stereoscopic viewing still is possible but the sense of depth is reversed).

Human stereoscopic viewing of the given stereo pair is not possible with the images stacked vertically.

Question 5. As we have seen, determining corresponding points in the left image and in the right image is the hardest part of stereo vision. A variety of things can go wrong in stereo matching. In a sentence or two for each, give a specific example of a scene where

(a) there are not enough locally distinct features that match

  Solution:
  Any scene containing extended smooth, featureless regions suffices. A stereo pair obtained by viewing a blank grey wall is one good example.

(b) there are too many locally distinct features that match

  Solution:
  Any scene containing many closely spaced, visually similar features suffices. A random dot stereogram is a canonical example.

(c) locally distinct features match incorrectly

  Solution:
  Any scene containing visually similar features that do not correspond to the same object point suffices. A surface, like that of a sphere, that curves smoothly away from view is the example cited in Forsyth & Ponce. In class, we added highlights/specularities on a smooth surface, like a sphere, as another example.

Hint: This is a question about the problem of stereo vision, not a question about the properties of any particular algorithm or technique used to do stereo matching.

Question 6. Lucas–Kanade estimate the 2-D motion, $[u, v]$, at a given point, $[x, y]$, in an image by computing the partial derivatives, I_x, I_y, I_t, in a window centered

at the given $[x, y]$. Their method assumes all points in the window are "inliers" with respect to the estimation of a single motion, $[u, v]$.

Suppose, instead, that there are multiple, distinct motions occuring within the window. Describe, in a few sentences, how you might use a Hough transform approach to detect and determine the multiple motions.

Solution: Consider a 2-D accumulator array whose parameters are u and v, respectively. In principle, both parameters can vary between $\pm\infty$. In practice, it's reasonable to limit both u and v to some finite range, say $\pm s$, where s is the maximum speed of motion we accommodate between successive frames. Divide the range $-s \leq u \leq s$, $-s \leq v \leq s$, into a finite number of bins.

As with standard Lucas–Kanade, compute I_x, I_y and I_t at each point in the window. For each point where one or both of I_x and I_y are non zero, consider the associated line, $I_x u + I_y v + I_t = 0$. In standard Hough fashion (for fitting data points to a line), cast votes for each u and v that lie on this line.

Once all points in the window have been considered, search the bins in the accumulator array to find clusters of votes (above a threshold). This determines all values of $[u, v]$ where there is sufficient support in the window for motion detection (according to the threshold selected).

4