✓ Congratulations! You passed!

Grade received 100%
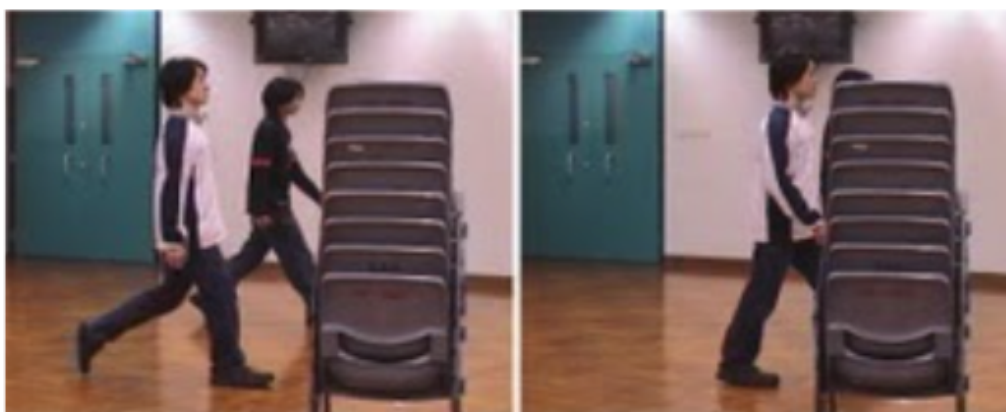
Latest Submission Grade 100%

To pass 70% or higher

Go to next item   Retake the assignment in **7h 59m**

1. The following images are sequential frames from a video. What would we expect the simple frame difference image between them to look like?   **1 / 1 point**

- ○ Fully black

- ○ Fully white

- ○ Only the men and chairs in white

- ◉ Only the men in white

  ✓ **Correct**
  A simple difference in frame only highlights parts of an image that change, with no recognition of foreground or background. Thus the men, as the only feature that changes between frames, would be the only highlighted objects in frame.

2. Which of the following about background modeling with a Gaussian Mixture   **1 / 1 point**

Model (GMM) is true?

◉ For each pixel, the method classifies it to be background or foreground depending on its location in the computed intensity histogram.

◯ For each pixel, the method calculates its largest change in intensity over time. Pixels with large changes are foreground pixels.

◯ For each pixel, the method calculates its smallest change in intensity over time. Pixels with small changes are foreground pixels.

◯ All the above

⊘ **Correct**
Background modeling with GMM first builds a normalized intensity histogram using the first n frames. For each subsequent frame, each individual pixel is classified as foreground or background depending on its location in the histogram. The first answer choice is correct.

3. In background modeling with a Gaussian Mixture Model (GMMs), σ refers to the standard deviation and ω refers to the scale of a Gaussian in the mixture. Which of the following properties does the Gaussian assumed to capture the foreground have?
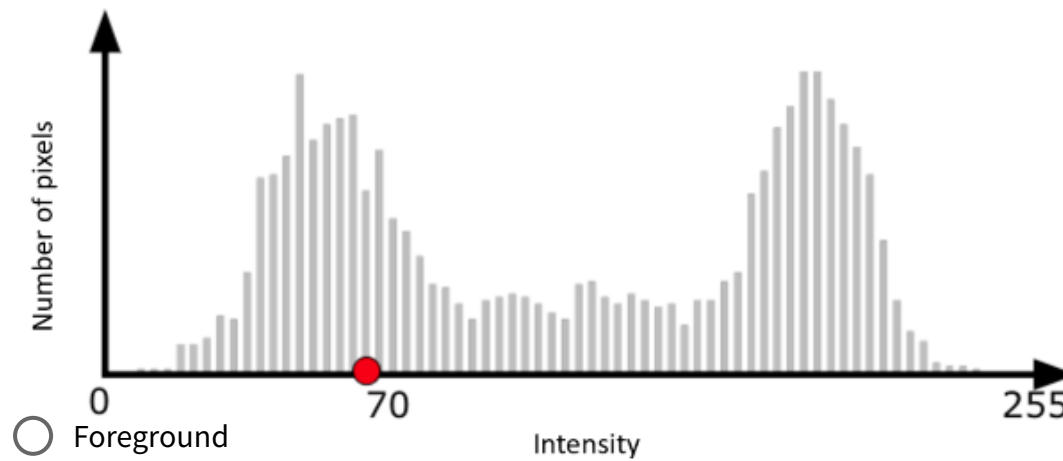
1 / 1 point

◯ $\omega \gg \sigma$

◉ $\omega \ll \sigma$

◯ $\omega = \sigma$

◯ None of the above

⊘ **Correct**
Small $\frac{\omega}{\sigma}$ corresponds to foreground pixels.

4. An intensity histogram for a pixel over time is shown below. At frame k, the intensity of the pixel is 70 (marked as the red dot). Is it more likely a foreground or a background pixel?

1 / 1 point

- ◯ Foreground
- ⦿ Background
- ◯ There is not enough information
- ◯ Noise

✓ **Correct**
The intensity more likely resides in the left-most Gaussian where $\frac{\omega}{\sigma}$ is relatively large. Therefore, it should be a background pixel at frame k.

5. Imagine a camera pointed at a screen that flashes between red, blue and white, in front of which people occasionally walk. We would expect the intensity distribution of a random pixel in this scene to be:

1 / 1 point

- ◯ Unimodal
- ◯ Bimodal
- ⦿ Trimodal
- ◯ Quadmodal

✓ **Correct**
The intensity distribution of a pixel will have modes at the intensities it picks up regularly. In this case, red, blue and white will all show up regularly for extended periods of time. Therefore, they will represent the high-frequencies (the peaks) in the intensity distribution. In contrast, the pedestrians will be rarer, and they will differ in appearance, thus, they will not result in a high-frequency event in the distribution.

6. In the following images, the red dots are the SIFT features corresponding to the object we are tracking, and the blue dots correspond to SIFT features outside the initial bounding box. Which of the following potential bounding boxes is most likely to result from running a feature detection-based method on the second image, initialized on the first?

1 / 1 point
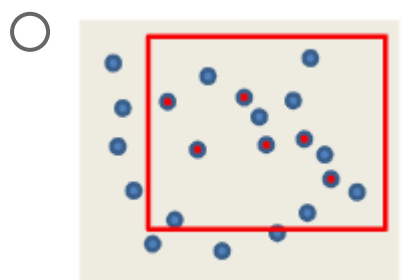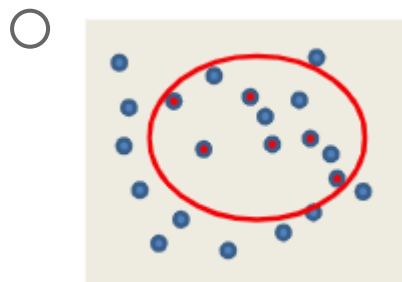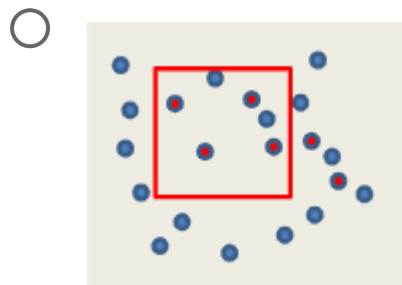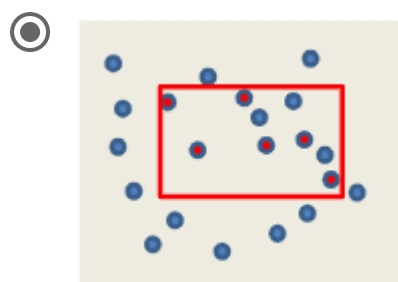
✓ Correct

Feature detection-based methods seek two goals when updating the bounding box: to maximize the number of inside points inside the new window, and to minimize distortion to the previous box. The first answer choice, while more or less undistorted, fails the first task, while the second and third answer choices are major distortions from the original box. Only the last answer choice both captures all the previous inside points and is close in shape and location to the previous box.

7. Assume we need to track the motion of a particularly shaped cloud in the sky. There are also other clouds with different shapes existing at the same time. The clouds are uniform in texture and color. Which of the following tracking methods is suitable in this case?

   2 / 2 points

   ● Template-based

   ○ Histogram-based

   ○ Feature detection-based

   ○ All of the above

   ✓ Correct

   Template-based tracking is the most suitable here, because we are discriminating on the basis of a concrete shape. With histogram-based tracking, we lose the location information, and since all the clouds are uniform in color, the clouds can become indistinguishable. Furthermore, we can't use Feature detection-based tracking, since the clouds are uniform in texture and color, lacking blobs of interest (SIFT feature candidates).

8. Assume we need to track the football in a game. The camera often zooms in to

   1 / 1 point

capture the football's landing location and zooms out to capture the entire football field. Moreover, the tracking system must run in real time on constrained hardware, rendering blob detection infeasible for example. Which of the following tracking methods is suitable in this case?

◯ Template-based

⦿ Histogram-based

◯ Feature detection-based

◯ All of the above

⊘ **Correct**
In this case, we need a tracking method that is robust to orientation (football rotating rapidly in the air), scale (camera zooming in and out), and computationally efficient. In addition, the football is unique in the environment and unlikely to be occluded. Therefore, a histogram-based method is suitable in this case.

9. Assume we need to track one of the dancers in the ballet show "Dance of the Little Swans," where all 4 dancers are wearing the same costumes. It is possible that the dancers might be partially occluded by the set props. Which of the following tracking methods is suitable in this case?

1 / 1 point

◯ Template-based

◯ Histogram-based

⦿ Feature detection-based

◯ All of the above

⊘ **Correct**
In this case, all 4 dancers look very similar to each other, so histogram-based methods will fail. Feature detection-based method is more suitable here because it can directly track the target dancer's face. Furthermore, since the dancers can be partially occluded by set props, a standard template-based matching could fail.

10. Which of the following real-world scenarios is object tracking usually applied to?

1 / 1 point

◯ Sports game

○ Shopping mall

○ Bank surveillance

◉ All of the above

⊘ **Correct**
Object tracking is needed for all the mentioned real-world scenarios.