|  |  |  |  |
| --- | --- | --- | --- |
| Question | Q1 | Q2 | Q3 |
| Marks | 6 | 8 | 6 |
| Total | 20 | | |

**Full Name (as in course registration): Apar Garg**

**Date: 15/10/2021**

**Course:** **Real Time Audio-Visual Sensing and Sense Making**

**Question 1.** For each question, select the **single** most appropriate answer, and **provide your justification** (no longer than two sentences).

a). Motion estimation method is used to estimate motion feature between two consecutive frames. Which of the following operation would **INCREASE** the dimension of the obtained motion features?

1. Apply the sub-pixel estimation, not the integer-pixel estimation, which is used in the block matching method.
2. Reduce the block size, which is used in the block matching method.
3. Increase the number of levels of the pyramid, which is used in the optical flow method.
4. All of above.

Option B

Reducing the block size would increase the no. of blocks for which motion features need to be obtained using matching/searching.

b). Which of the following proposals is **feasible** for recognizing a single person dancing in an indoor dancing studio?

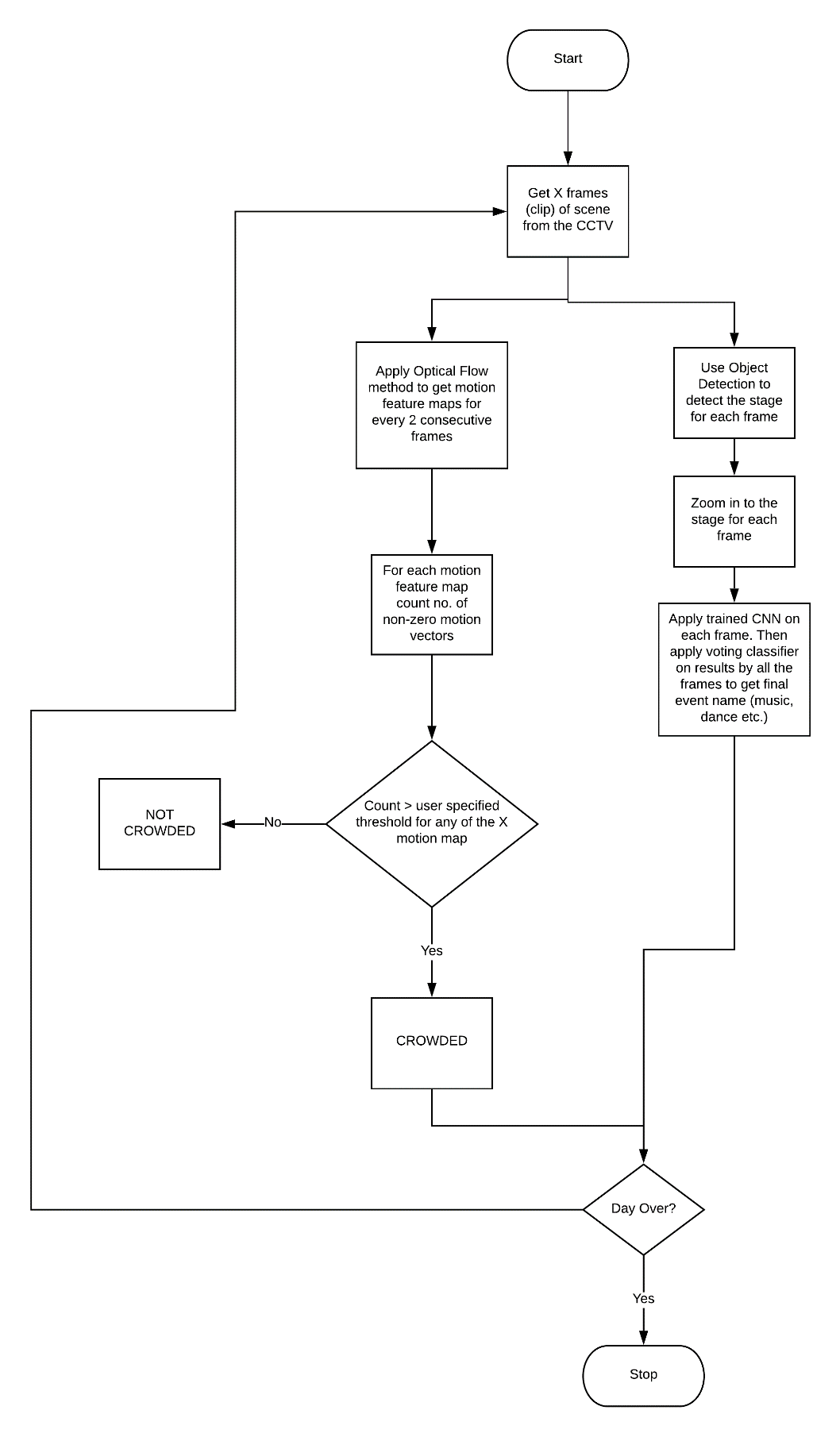
1. Generate a motion history image from the dancing sequence, followed by an image classification approach.
2. Estimate motion field for every two consecutive frames, followed by a sequence classification approach.
3. Extract human body shape from each individual image, followed by a sequence classification approach.
4. All of above.

Option A

The motion history map has shape information and also the sequence information obtained from multiple frames. Classification is performed one time (on the motion map).

Actually, all these three methods are suitable for this indoor scenario.

**Question 2.** In public venues, crowd size is a key indicator of crowd safety and stability. Monitoring the people number and crowd density levels are important. You are engaged by a security surveillance company to develop a security solution for New Year countdown party at the Sentosa beach, as illustrated in Figure below. The proposed system is required to have the following two functionalities: (a) human counting (crowd density); (b) stage event classification (sing, dance, etc). The camera is mounted at the entrance facing the stage (a snapshot of the CCTV image is illustrated in the right image). You can draw a flow chart to justify your proposed system.

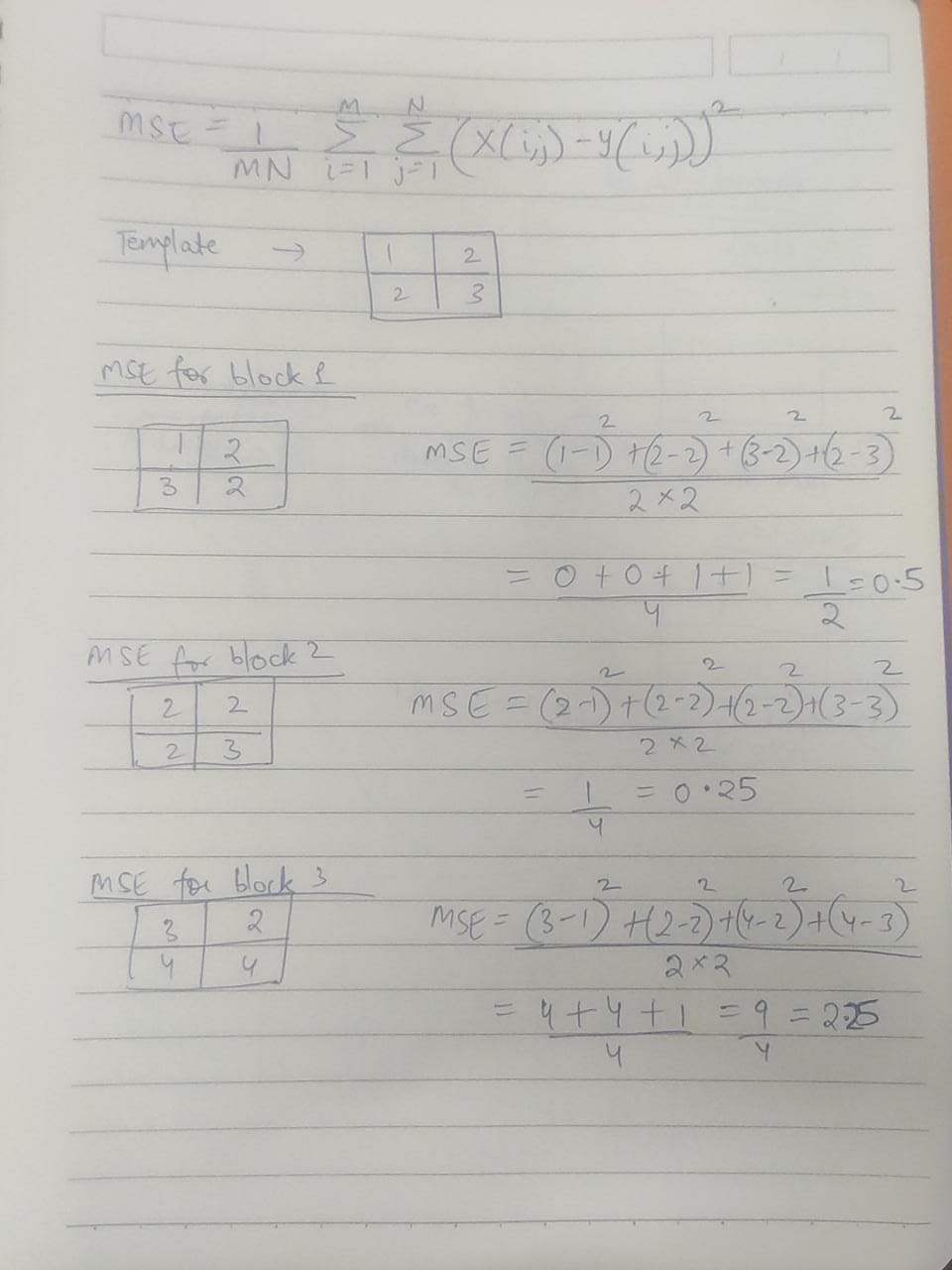


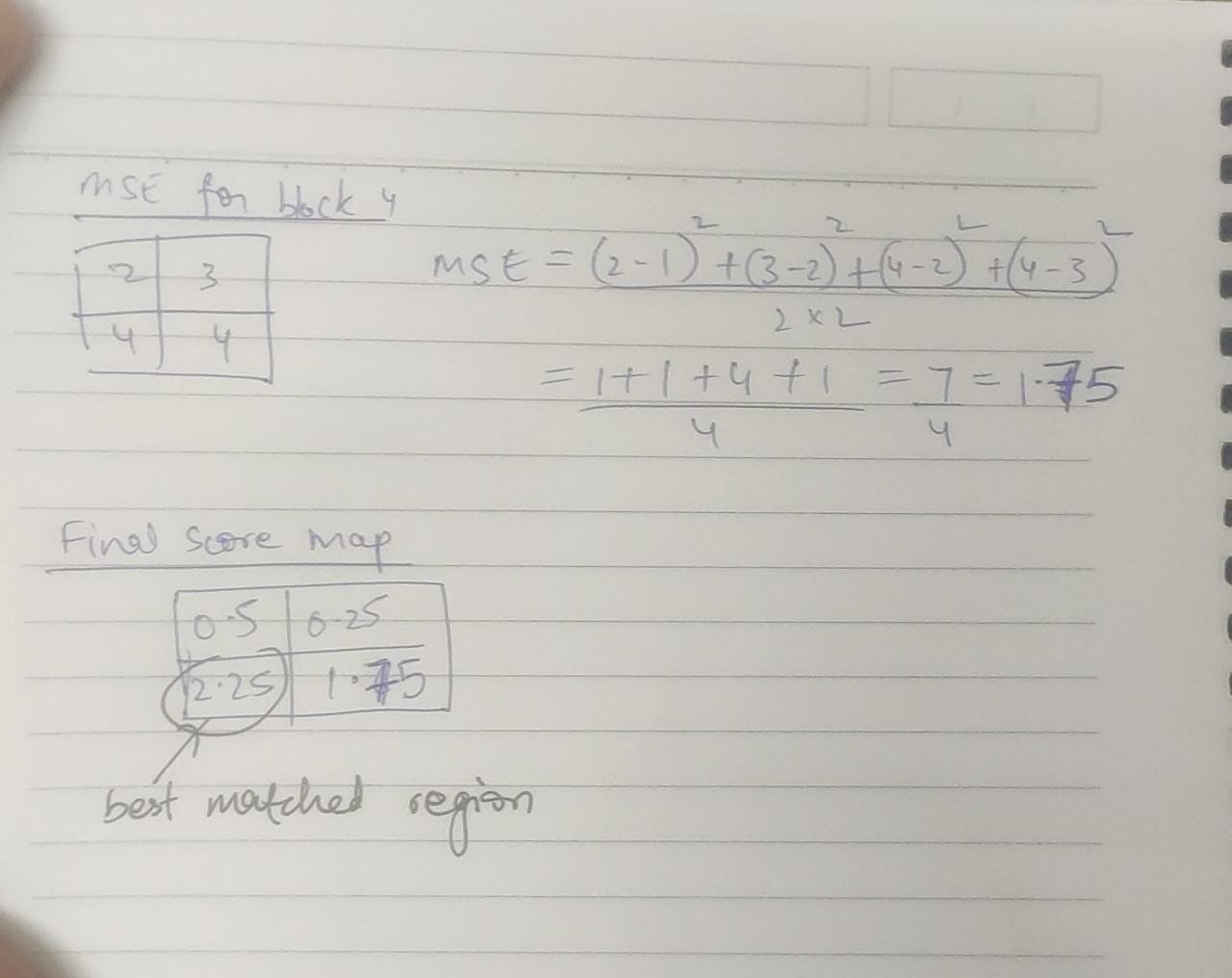
**The no. of motion vectors might not be accurate enough for a crowd classification.**

**Stage region can be obtained using the predefined ROI.**

**Question 3.** You are engaged by an electronic entertainment company to develop a human tracking system. Given the search reference image in Table 1, and the object template illustrated in Table 2, apply the *mean square error* (MSE) method to determine the best matched image region (integer pixel accuracy) in the search reference image. Show your calculations to justify your answer.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1 | 2 | 2 |  | 1 | 2 |
| 3 | 2 | 3 |  | 2 | 3 |
| 4 | 4 | 4 |  | Table 2. The object template. The numbers are gray-scale intensity values. | |
| Table 1. The search reference image. The numbers are gray-scale intensity values. | | |  |





Will you use the largest MSE (2.25) or smallest MSE (0.25) to find the best matching region?