

Homework 4:

Introduction to Image and Video Processing

Huang Yiping

1 Motion Energy Images

- 1.1 Finding the MEIs in the activities videos you chose (Matlab or Python). Make sure you choose w to be a reasonable temporal window (there is no one correct value for w). Display the resulting MEIs for two subsequences of w frames. Use what you consider are reasonable values of w , and explain why you chose them.**

Answer:

Motion Energy Images(MEI) is defined as $E_{\tau}(x, y, t) = \cup_{i=0}^{\tau-1} D(x, y, t - i)$, where τ is the window size, and $D(x, y, t)$ is a binary image indicating whether there is motion under the (x, y) point in the time t . To determine this value, I use the magnitude of the interframe difference, i.e., if $|I(x, y, t) - I(x, y, t - 1)| > \epsilon$, then I set $D(x, y, t)$ to 1. The $I(x, y, t)$ is the intensity at (x, y) in the t point of time, and the ϵ is a small value to remove the background noise due to the sensors of the camera, the value of pixels are transformed to double and I set the default value of ϵ to be 0.05. All figures below are produced with this value if without further explanation.

A window of size 5 is chosen initially, and the sequence images are shown below for the jump, bend and skip action from the Weizmann actions dataset.

The window size of 5 is chosen somewhat subjectively. However, since the jump action is repeated 3 times in the video, and in total there are 38 frames in the video, so in average, one jump action takes about $38/3 = 12$ frames to complete. With 5 sequence in display and with each frame captures the previous 5 frames of movement, we can expect the displayed MEI sequences are relatively complete in describing the motion.

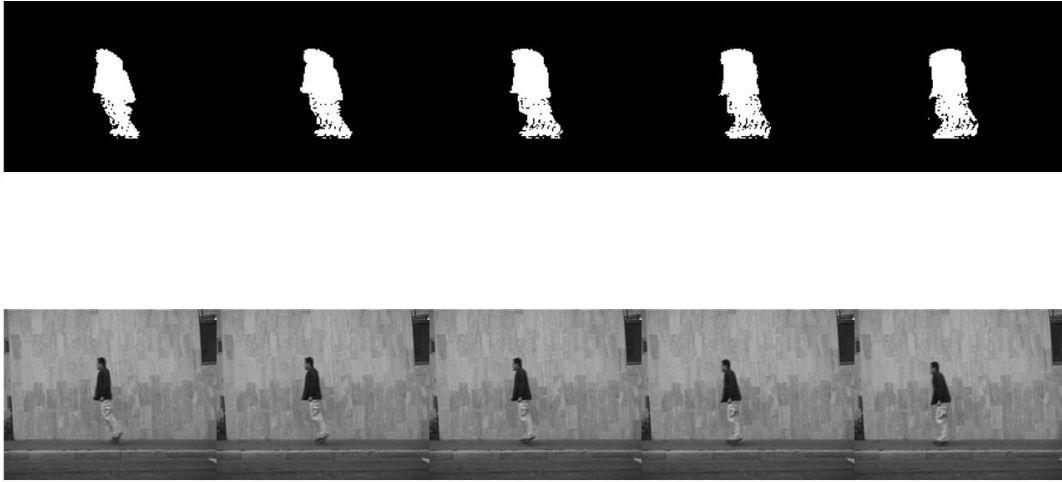


Figure 1: *MEI sequence of shahar jump using window of size 5 and the corresponding frames starting from the 20th frame*

The bending sequence is plotted similarly as shown below. We observe that for bending action, 5 sequences of MEI with a window size of 5 are not enough to describe the action completely.

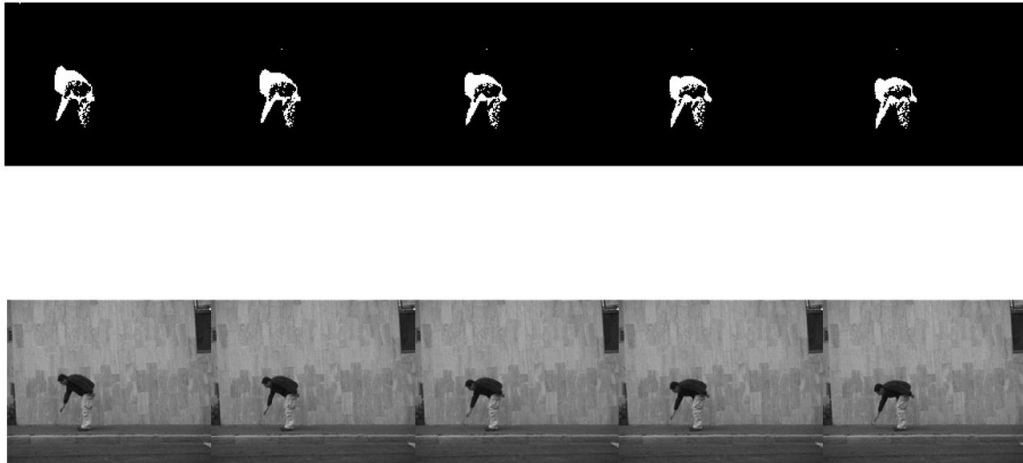


Figure 2: *MEI sequence of shahar bend using window of size 5 and the corresponding frames*

Finally, the skipping sequence is shown as well. We see that a sequence of 5 frames with a window size of 5 is enough to describe the skip action.



Figure 3: *MEI sequence of shahar skip using window of size 5 and the corresponding frames*

Seeing that the bending action is not complete, a window size of 10 is used and a plot of 10 sequences are shown for each of the action similarly as above figures.

From the blobs, we see that the difference in MEI sequence between the frames is further reduced due to the bigger size of the window. The bending action is now much more complete than in the previous figure, although still not fully. However in calculating the action moment, the whole series of video are taken into consideration, so it is ok that the sequences shown here are not fully complete.

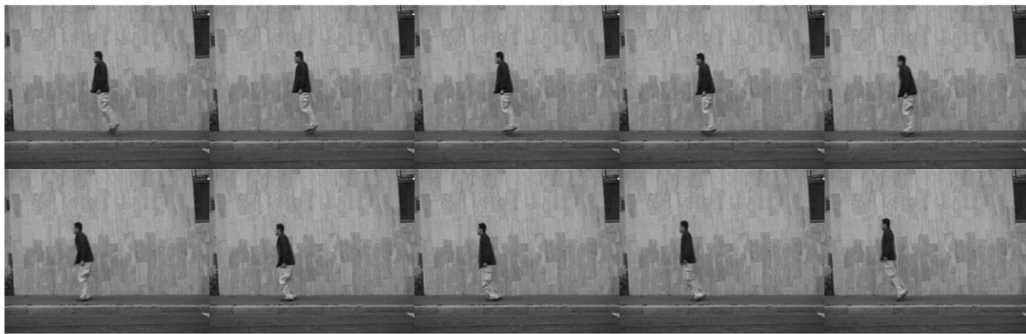
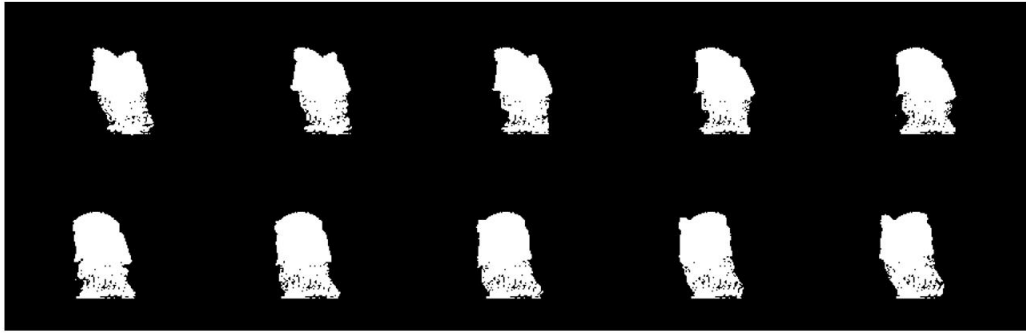


Figure 4: *MEI sequence of shahar jump using window of size 10 and the corresponding frames*

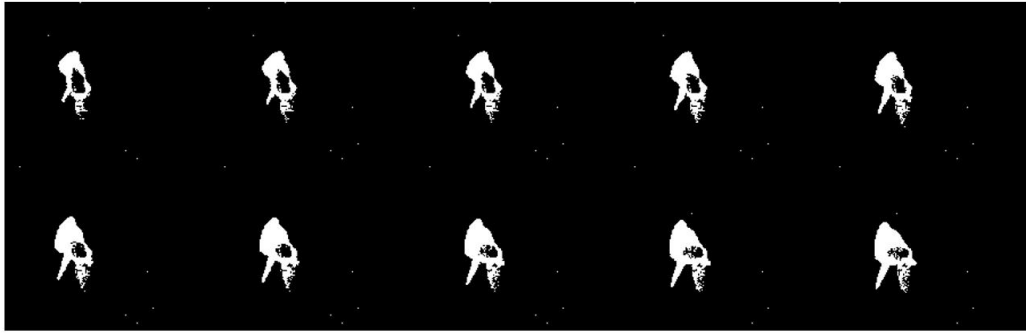


Figure 5: *MEI sequence of shahar bend using window of size 10 starting from the 15th frame and its corresponding frames*

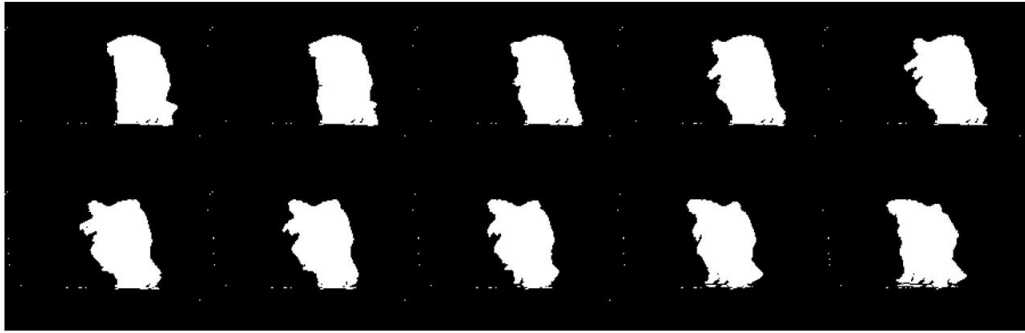


Figure 6: *MEI sequence of shahar skip using window of size 10 and its corresponding frames*

1.2 Bonus

1.2.1 Add noise to the image frames

Answer:

For demonstration purpose, Shahar jump video is chosen an example to add noise on. A Gaussian noise with variance 0.1 is added to each frame of the video using *imnoise* function.

1.2.2 Then find MEI (using the magnitude of inter-frame differences or optical flow)

Answer:

For convenience sake, figure 1a is shown once again for comparison with the noise sequence. The same method of measuring magnitude of inter-frame differences is used to calculate the noise sequence. Clearly, due to the randomness of noise, much of the background has also been taken as the moving object and thus the object in moving is hard to tell from the background.



Figure 7: Sequence of shahar jump using window of size 5

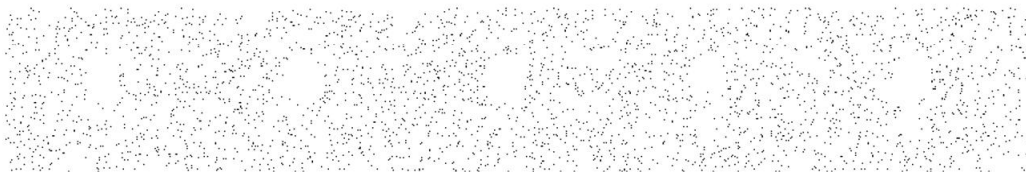


Figure 8: Sequence of shahar jump using window of size 5 with added gaussian noise

1.2.3 Apply de-noising (you decide what kind) and find the MEI again. Display, and discuss/compare the results before and after de-noising

Answer:

To remove the noise, a Gaussian average filter with size 5×5 and σ of 10 is used to filter the video, and a bigger threshold value of 0.14 is used to compute the MEI. The result sequence is shown below. As can be seen, much of the background noise has been removed, but the leg movement is harder to see. This is because the background runs through the legs and the size of leg is relatively small for the filter compared to the torso so these movements are not salient anymore when energy is later counted.



Figure 9: *denoised sequence of shahar jump using window of size 5*

1.3 Clean up the MEIs (binary images) with one or more morphological operations. Explain why you choose these operations. Display and discuss the results

Answer:

The close process is applied to the MEI. The result applied with the structure element of a disk of radius 3 is shown below. The process is chosen since close operation is dilation followed by erosion. Thus it can fill the remaining gaps in the object and also smooth the contour of the object, which helps with image segmentation.



Figure 10: *application of close operation in the MEI sequence of shahar jump using window of size 5*

1.4 Find the outline of the MEI using a method of your liking, such as edge detection, morphological boundary extraction. Display it

Answer:

The erosion operation is chosen to find the outline of MEI. The erosion is applied with a disk shape element with radius of 3. This outline is produced by the difference of the original image and the eroded image.



Figure 11: *outline by method of erosion of MEI sequence of shahar jump*

1.5 Extract the shape descriptor for the MEI outlines of the actions using Hu or Zernike moments, using ready-made functions

Answer:

Hu's moments are formed by a vector of 7 components, where each one is formed by the combinations of central moments up to order 3. They have good properties, namely that they are translation, rotation and scale invariant, which means they can be used to extract unique features from the images.

Here the outlines of MEI sequences of Shahar jump with a window size of 5 is chosen to compute the Hu's moment. To get the Hu's moment for the outlines, the Hu's 7 moments for each frame is computed, then an average is taken across the frames to get the final averaged vector for the video. The results are put in the matrix M as shown below, where the first row corresponds to the row vector from the jump action, the second row to the bend action and the third row to the skip action.

$$M = \begin{pmatrix} 0.5700 & 0.1795 & 0.0377 & 0.0745 & 0.0033 & 0.0294 & 0.0002 \\ 0.5357 & 0.0765 & 0.0695 & 0.0488 & -0.0506 & -0.0118 & 0.2662 \\ 0.4817 & 0.1945 & 0.0548 & 0.0511 & 0.1049 & 0.0014 & 0.0226 \end{pmatrix}$$

1.6 Do a simple comparison of the shape descriptors you found between the three actions (e.g. by finding the Mean Squared Error or any other difference measure between the Hu descriptors of the three MEIs etc). Show and discuss your results

Answer:

To compare the differences between them, the pairwise euclidean distances are calculated, which is defined as $\sqrt{\sum_{i=1}^7 (x_i - y_i)^2}$. The result of each pair is then put in another matrix D as shown below, where D_{ij} indicates the distance between the i th row and j th row of matrix M .

$$D = \begin{pmatrix} 0 & 0.2980 & 0.1431 \\ 0.2980 & 0 & 0.3174 \\ 0.1431 & 0.3174 & 0 \end{pmatrix}$$

We can see that D_{13} has the smallest value, i.e., the action of jump is most similar to that of skip among all three actions considered, which is intuitive from our daily experience. On the other hand, the most dissimilar action pair is ones between bend and skip. These results indicate that we can use the Hu's moment to detect different actions from each other.
