Walk This Way: A Study of the Eigenspace Transform as a Motion Discriminator

Z Cai & J Provine, Cornell University

for Dr. Reeves' Machine Vision Class

12.04.2000

Abstract

This project studied the usefulness of the Eigenspace Transform (EST) in a motion discrimation algorithm. The data was videotaped motion of various people performing three distinct motions: walking, hopping, and leaping. The videotaped data was converted into binary VisionX images. An algorithm for performing the EST was written and used to project the data into eigenspace. A preliminary study of the discrimination of the different motions in eigenspace was performed and revealed promising motion discrimination based within three dimensional eigenspace.

1 Introduction

One common machine vision application is to teach a computer to discriminate some dataset automatically, to save the man-hours or boredom attributed to these tasks. This project began by an interest into the possibility of human identification from gait patterns[1][2]. To simplify to a more feasible half-semester project, the task shifted to discriminating between different human motions: walking, hopping, and leaping. Despite this switch the methods explored in this project will be of use in many automatic motion analysis techniques, which has many applications in surveilance (alert guard if unknown character is walking about premises) and medical fields (analysis of walking patterns to determine joint disfunction).

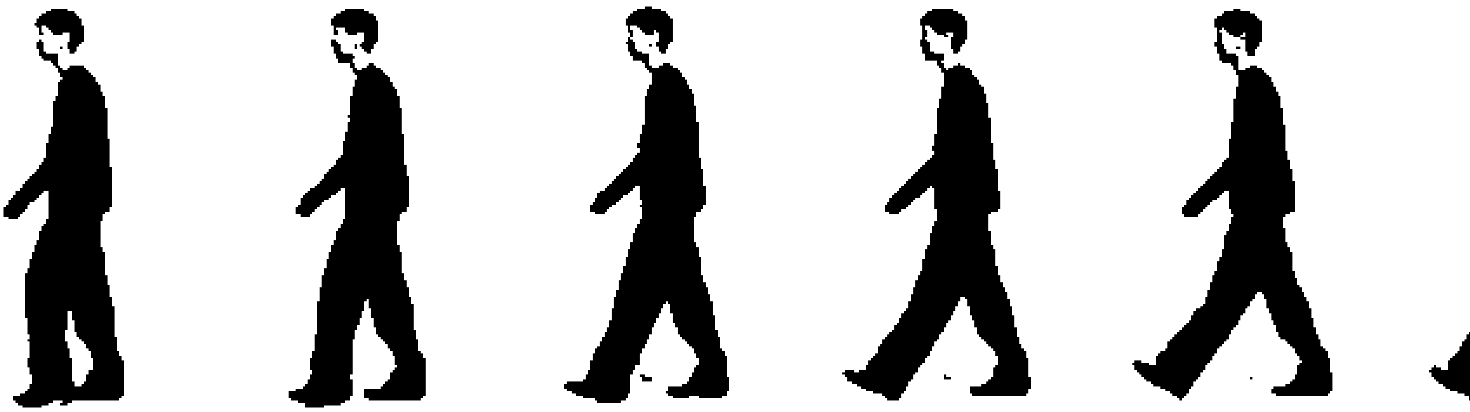

Figure 1: A Prime Example of Walking

Walking, hopping, and leaping were chosen because of their ease to discriminate visually, creating a supposed ease of discrimination by machine vision methods. Additionally, the motions could be rigorously defined. Namely, walking is ambulation at a normal speed and motion (think runway model), hopping is short continuous boncing along in one direction (think bunny), and leaping is jumping without strenuous effort from a standing start (think kangaroo). Figures 1, 2, and 3 provide examples of walking, hopping, and leaping data.

The EST was chosen as the desired method for automating the motion discrimination problem primarily because of the readily available literature [1][2]on this method, but additionally because the EST was a new topic not discussed in class. Utilization of the EST required formatting our images in a special way. The format and methods are discussed in Section 2. Then in Section 3 the EST is rigorously defined and explanation is given as to our implementation. Section 4 describes the tests were performed and evaluates the preliminary results achieved. Finally, in Section 5, the various avenues for future efforts on this project are discussed.

Figure 2: A Prime Example of Hopping

2 Binary Image Extraction: Our World in Black and White

The binary image extraction process consists of six stages. First, the video tape recording of the different motions was converted into digitized image sequences compatible with VisionX. Adobe Premier converts the digitized video into a .tif format, which is subsequently reformatted into a VisionX image sequence. Second, the fringe caused by video interlacing was removed. Video interlacing is the mixing of even pixel rows from a frame with the odd pixel rows of the preceding frame. This phenomenon can be easily remedied by replacing the pixel values in every other image row with the average of the two vertically adjacent pixel values. Third, a threshold value was chosen to create a binary image. It was determined that a threshold value of 65 from a 0 to 255 scale worked most effectively for the data set. Fourth, the binary thresholded sequences were passed through a 3X3 median filter to remove "salt & pepper" noise.

Fifth, to normalize the figures, a 455X250 window was centered around the region's center of mass. Normalization was performed to focus the motion discrimination on the subject's physical form and not the subject's relation to the fixed background. The dimensions of the window were chosen to contain the entire body throughout the sequence. Sixth, the windowed sequences were downsampled by a factor of four to 227x125. This was achieved by sampling pixels from every other column and row. The reason for downsampling is to lower calculation time and appease the god of Graduate Disk Space Allotment. Please refer to the attached *zjprogram* for the source code of this algorithm.

Figure 3: A Prime Example of Leaping

3 The Eigenspace Transform: It's Eigenriffic

The impetus behind transforming the image data to another state for analysis is to compress much of the information and evaluate only those most important features, thereby increasing efficiency and necessary memory for the discrimination. The method chosen was the Eigenspace Transform (EST). For each motion class, the covariance matrix of each frame is computed:

$$Q_h(t) = \sum_{i=1}^{N}(h_i(t) - \bar{h}(t))(h_i(t) - \bar{h}(t))^T \tag{1}$$

where $Q_h(t)$ is the covariance matrix of hopping for frame t, N is the total number of versions of the motion, $h_i(t)$ is the i^{th} version of hopping for frame t, and

$$\bar{h}(t) = \frac{1}{N}\sum_{i=1}^{N} h_i(t) \tag{2}$$

is the mean of the hopping versions. Similar covariance matrices are found for walking and leaping.

The eigenvalues of the covariance matrices are calculated. Only the k largest eigenvalues are retained and these comprise our feature vector. Each 227x125 image frame is thus reduced to a single point in k-dimensional eigenspace; and an image sequence maps to a trajectory through eigenspace. Overall the size reduction is from 227x125x30 in the image spacio-temporal domain to kx30 in the eigenspace.

Another advantage of the EST is that by orthogonalizing the basis of the covariance matrix, it returns the features of maximum variation in the data set. This quality has a somewhat mystical effect because it is not known what these features of maximum variation

Figure 4: An Example of the Input Digitized Images

are. It could be the feature of largest variation is not very adept at the discrimination sought. Most often the value of the EST cannot be approximated *a priori*.

4 Tests & Results: Failure or Success?

Using N=5, three sets of 30 covariance matrices were computed for the respective motions. The feature vector was then defined by the k=3 largest eigenvalues (e3, e2, e1) of each covariance matrix. Hence, each motion maps to a length-30 path in 3-dimensional eigenspace. From the data, each path occupies a recognizably distinct location in the three dimensional eigenspace. Figure 6 displays the trajectory in 3-dimensional eigenspace for the leaping training data set.

The test portion of this project consists of discerning the motion portrayed in a new image sequence by applying the nearest neighbour algorithm with the trained data sets. First, three sets of 30 covariance matrices from the new image sequence are computed, testing the new dataset verses each of the motions. Each set is calculated using the following function (testing verses hop for example):

$$Q_h(t) = \sum_{i=1}^{N} (x(t) - \bar{h}(t))(x(t) - \bar{h}(t))^T \tag{3}$$

where $x(t)$ is the unknown image sequence. It should be noted that $x(t)$ in the case presented was separate from those image sequences used to train and create the eigenspace. However, the subject of the unknown image sequence was utilized in creation of the training set as well.

Once more, the three largest eigenvalues of each covariance matrix were taken to create three length-30 paths in eigenspace corresponding to the three motions. The point-by-point

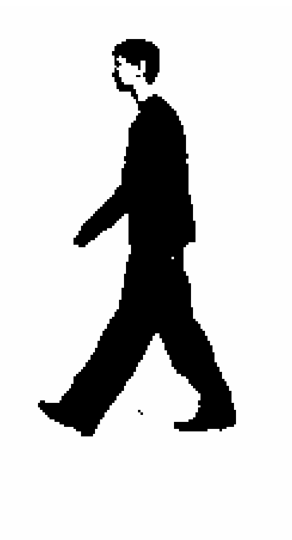

Figure 5: An Example of the Output Processed Binary Images

euclidean distances between the three experimental paths and the three corresponding trained paths were calculated. Subsequently, the mean of these distances was obtained. The nearest neighbour algorithm then selected the minimum value of the three means which should correlate with the discerned motion.

The aforementioned test was implemented to recognize an untrained hop sequence. The mean for hopping was the smallest, indicating a correct match. This mean was approximately an order of magnitude smaller than the means of the unmatched paths. Success!

5 Conclusions & Future Work: Much more exciting stuff

The EST provided a successful method for discriminating the human motions walking, hopping, and leaping. Although the training process was computationally extensive and not very fast, the overall data reduction for discrimination provides a potentially efficient method for discrimination once the system has been trained.

Future possibilities for this project are numerous and look promising. Further tests with more varied and noisier test image sequences to be discriminated is a logical next step. Adding more eigenvalues to the analysis would be another interesting step to consider, and cursory viewing of these values indicates that this could further discriminate the motions. Additionally, a comparison can done between this method and other methods of producing feature vectors (Fourier Discriptors, method of Moments, etc.).

The authors found this project to be instructional and delightfully challenging. Numerous topics were considered for pretty much the first time (namely, the EST, Numerical Recipes in C packaged functions, html coding, xv image system, and many of the VisionX commands and capabilities). Overall this was an excellent learning experience which presented many problems and resulted in a truly enjoyable outcome.

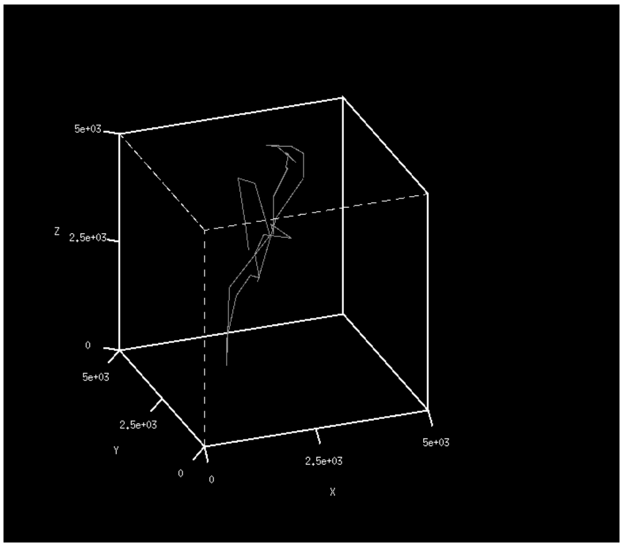

Figure 6: An Example of the Trajectory in 3D Eigenspace for Leaping

6 Acknowledgements

The authors would like to thank their TA Andinet Asmamaw for his invaluable help in all phases of this report. Dr. Anthony Reeves was also helpful in determining the scope of the project and aided in our understanding of the EST. Dan McCormick was helpful in preparation of image files for the report. Finally, the authors wish to thank netflix.com for their necessary input.

References

[1] H. Murase, R. Sakai, *Moving object recognition in eigenspace representation: gait analysis and lip reading*, *Patter Recognition Letters* 17, (1996).

[2] P.S. Huang, C.J. Jarris, M.S. Nixon, *Human gait recognition in canonical space using temporal templates*, *IEEE Proceedings* online no. 19990187 (1999).

[3] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge Univeristy Press, New York (1991).

[4] R. Jain, R. Kasturi, B.G. Schunk, *Machine Vision*, McGraw-Hill, New York (1995).