Individual Recognition from Periodic Activity Using Hidden Markov Models

Qiang He and Chris Debrunner
Colorado School of Mines
Division of Engineering
{qhe, cdebrunn}@mines.edu

Abstract

We present a method for recognizing individuals from their walking and running gait. The method is based on Hu moments of the motion segmentation in each frame. Periodicity is detected in such a sequence of feature vectors by minimizing the sum of squared differences, and the individual is recognized from the feature vector sequence using hidden Markov models. Comparisons are made to earlier periodicity detection approaches and to earlier individual recognition approaches. Experiments show the successful recognition of individuals (and their gait) in frontoparallel sequences.

1. Introduction

In automated visual surveillance systems, recognition of humans and their activities is generally the most important task. Video retrieval systems also can also use such capabilities to expand the range of queries they can handle. Two forms of human recognition can be useful in these contexts: the determination that an object is from the class of humans (which we call human recognition), and determination that an object is a particular individual from this class (which we call individual recognition). In this paper we focus on the latter problem, but we expect the approach to be equally applicable to the former. In addition, the approach can successfully distinguish periodic activities.

Human action can be either periodic or non-periodic, and in our approach we processes these two cases differently. In this paper, we describe a technique for recognizing individuals from periodic motions, specifically walking and running. Periodicity is an important component of the information in an activity, so we emphasize periodicity detection in our approach. Figure 1 shows three frames from an example sequence processed by our approach.

Our approach consists of the following steps. We first segment the image sequence based on motion, and we compute the Hu moments of segmented motion regions. We then match Hu moments over time to determine the degree of periodicity and the period of the motion. Then hidden Markov models (HMMs) are used to recognize the individuals from sequences of quantized Hu moment vectors. The periodicity information is used to compute the size of codebook, to compute the number of states and to distinguish the two kinds of activities (walking and running).

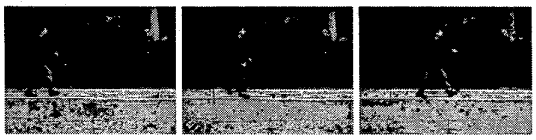


Figure 1 Three frames from an example of a walking sequence.

2. Previous work

The works most related to ours are Yamato *et al.* [1]* and Little and Boyd [9]. Yamato *et al.* made use of hidden Markov models to recognize the human actions based on low-resolution image intensity patterns in each frame. These patterns were passed to a vector quantizer, and the resulting symbol sequence was recognized using a HMM. Their method did not consider the periodicity information,

* Many other methods make use of HMMs for gesture or action recognition [2-8], but Yamato *et al.* use simple features more similar to the Hu moments we use.

0-7695-0939-8/00 $10.00 © 2000 IEEE

and they also have no systematic method for determining the parameters of the vector quantization.

Our approach differs in that it explicitly models and uses the periodicity, and in that it uses features based only on the moving portion of the image (background pixels are not used in the feature computation).

Davis and Bobick [10] used Hu moments of motion history images (MHI) and motion energy images (MEI) computed over a short time period from the foreground regions in each frame. These images (and their Hu moments) are distinctive for a variety of actions, and can therefore be used for action or gesture recognition. Davis and Bobick recognize the extracted Hu moment vectors by matching them to trained Gausssian distributions in the features space using the Mahalanobis distance. As in our approach, these features are only invariant to viewpoint to the extent that the Hu moments provide invariance to rotation, translation, reflection, and scale. This work focused on non-periodic actions and hence did not consider the periodicity information.

Little and Boyd [9] compute a set of moment features of the optic flow. The time sequence of each feature is then analyzed for periodicity and the phase difference between features is computed. These phase differences are the quantities used for recognizing individuals. They demonstrate high recognition rates in a large data set collected under very controlled circumstances. Our approach recognizes motion segmentations instead of optic flow, which we feel is a more stable set of features. Our method achieves recognition rates similar to Little and Boyd over a smaller data set, but one with larger variations in the collection environment.

Detection and estimation of image sequence periodicity has been more thoroughly explored in previous work than individual recognition. Generally speaking, there are two types of methods for periodicity detection: those based on the Fourier transform and those implemented in the time domain. Typically, a stochastic signal can be decomposed into the deterministic (periodic) component and the non-deterministic (random) component after Fourier transformation [11]. In frequency domain, the deterministic component corresponds to the harmonic peaks and the non-deterministic component corresponds to the smooth part of the spectrum. Therefore, the energy of the spectral harmonic peaks is a good measure of the periodicity. In Polona's method [12], a periodicity measure based on 1-D Fourier transforms along the temporal dimension is computed. The periodicity measure is defined for each pixel as the normalized difference of the sum of the power spectrum values at the highest amplitude frequency and its multiples, and the sum of the power spectrum values at the frequencies halfway between. If the ratio is close to one, the motion is periodic, otherwise it is not periodic. The periodicity for the entire image is defined as the maximal value of the average value of each special periodicity measure. Tsai makes use of the Fourier transform of the autocorrelation of the curvature of a spatio-temporal curve to study cyclic motion [13], and uses a median filter to estimate the background. Liu and Picard [11] measure the amount of periodicity in a tracked object pixel using the *temporal harmonic energy ratio*, which is the ratio of the energy in the spectral peaks to the total energy in the spectrum.

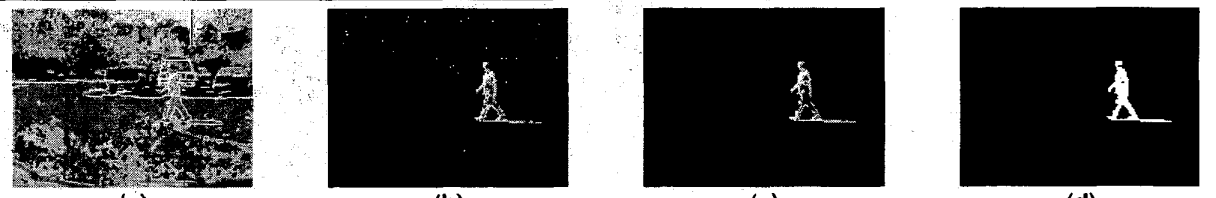
The second type of method implements the periodicity analysis in time domain. Little and Boyd [9] used maximum entropy spectrum estimation to find the period. Allmen and Dyer [14] used curvature as a low-level description of motion, and its scale-space as a representation. An advantage of curvature scale-space is that it is possible to detect cycles at any scale. Our method computes a minimum in the feature vector difference, and is reliable across our entire dataset.

3. Algorithm description

The training phase and the recognition phase of our approach make use of the following processing steps: motion segmentation, feature extraction, periodicity detection, feature quantization, and HMM recognition. In the training phase, motion segmentation is performed on each frame and features are extracted. The periodicity is determined from the features, and the number of frames in the period is used as the number of states in the HMM. One codebook and one HMM is trained for each gait of each individual. The training sequences are used to generate the codebook for the vector quantization [15]. The training sequences are converted into symbol sequences using this codebook, and the HMM is trained from theses symbol sequences.

In the recognition phase, motion segmentation is performed on each frame and features are extracted. The periodicity is determined and is used to determine the gait. For each trained HMM of the selected gait, the features are quantized into symbol sequences, and the probability that the HMM produced the symbol sequence is computed (see the forward algorithm in [16]). The output of the recognizer is the individual and gait of the HMM with the maximum probability.

The following subsections describe the processing steps in more detail.



(a) (b) (c) (d)

Figure 2 The steps of the motion segmentation process. (a) shows the original image, (b) shows the pixels above the change threshold, (c) shows the image of (b) after median filtering, and (d) shows the resulting binarized segmentation. Note that the shadow is detected too, which causes failures in our activity recognition approach (as well as in others).

3.1. Motion segmentation

In order to study the cyclic human action, we need to segment the objects from the background. This is accomplished by robustly estimating the statistics (mean) of the background and segmenting any pixels that do not fit the statistics. Median filtering along the temporal direction is performed for each pixel to estimate the mean of its background value. Because of wind and unstable light sources, small blobs always appear. We make use of spatial median filtering in each segmented image to remove the small blobs. The Figure 2 is an example of typical segmentation results.

3.2. Feature extraction

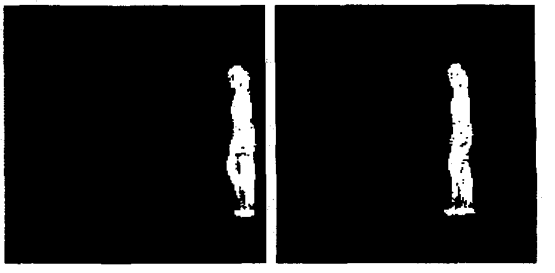
Hu moments as described in [17] and [18] remain constant under translation, rotation, reflection, and scaling, so they are good descriptions for image segmentations computed using the algorithm described in the previous section. We use the Hu moments of the segmented images as the image feature vectors.

3.3. Periodicity detection

The periodicity detection is based on the similarity for Hu moments,

$$N_t = \arg\min_{i \in S} | H_t - H_{t+i} |^2 ,$$

where N_t is the number of image frames in one cycle starting at frame t. H_t is the Hu moment vector at time t, S is the search range. The time interval between two contiguous image frames is known, allowing us to compute the periodicity from N_t. However, the actual periodicity is double the computed periodicity because of the symmetry of walking and running in the side-views that we use. This method is simple, and experiments show that it is more exact and more reliable than the Fourier transformation methods we tested. At the same time, tracking of complex features is unnecessary. As is common in periodicity estimation, we use the average value of the periodicities computed over several cycles as the overall periodicity of each activity. Figure 3 shows the two matching patterns detected using this method.



Frame 14 **Frame 31**

Figure 3 Examples of the matching frames selected by the Hu moment matching method of periodicity estimation.

3.4. Feature quantization

The Hu moment feature vectors are quantized to provide a discrete set of symbols as input to the HMM. We generate the codebook by averaging the corresponding feature vectors from several successive cycles of the training sequence. The codebook size we use is the number of frames in a period. Given the codebook, input feature vectors can be converted to symbols by finding the closest codebook vector [15]. The symbols corresponding to the closest codebook vectors are passed as input to the HMMs.

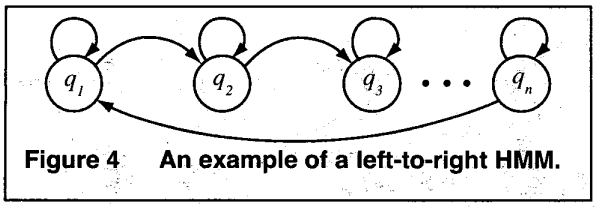


Figure 4 An example of a left-to-right HMM.

3.5. Hidden Markov models

It has been shown that hidden Markov models are successful tools for modeling and classifying dynamic behaviors. For example, HMMs have been widely and successfully applied to speech recognition. There are many similarities between speech recognition and human action recognition. A human being's voice and human action are both dynamic behaviors. Each person has characteristic speech and action patterns. If the characteristic features of a person's speech or actions are extracted, they can be recognized by a HMM.

In order to model human gait, we make use of left-to-right discrete HMM (Figure 4.), where one state of the HMM represents one gait state. We select half the number of frames in one cycle as the number of states in the HMM. For our left-to-right discrete HMM, there are two possible transitions in one HMM state, one to the same state and one to the next state.

The HMM approach to individual recognition is actually a classification method, that is:

$$M = \arg\max_{j=1,2,\ldots,C} P(M_j \mid S)$$

where S is the input symbol sequence coming from the vector quantization of the motion feature vectors, M_j is the j^{th} HMM, and C is the number of HMMs.

HMM parameter estimation makes use of an Expectation Maximization (EM) approach called the Baum-Welch algorithm. The HMMs are trained on the symbol sequences generated from the vector quantized Hu moment vectors computed from the motion segmentations of the training sequences.

4. Experiments

The experiments use image sequences of three individuals (who we will refer to as A, B, C) in both walking and running gaits. We assume that there exists one constant viewpoint and the human activities are frontoparallel to the camera and with constant speed within a particular sequence. The image sequences were collected using a Cannon Optura miniDV camera, were captured to a computer at 30 frames per second with 640 by 480 resolution, and were converted to grayscale before

Figure 5 Walking sequence segmentations.

Figure 6 Running sequence segmentations.

motion segmentation. There are ten sequences for each individual in each activity. Each sequence contained many cycles of the activity. Two of each of these ten are used to train a HMM and eight are used to test the HMM. For an input observation sequence, we set a periodicity threshold of 24 frame/cycle to distinguish walking and running. The ten sequences for a particular individual and activity had significant variation in backgrounds, walking and running speeds, directions of motion (left or right), distances to the camera, and attire of the individuals.

Unlike the experiments of Little and Boyd [9], some of the sequences were collected in lighting that produced shadows. To eliminate this effect, these sequences were manually edited to remove the shadows. An alternative would be to detect shadows based on chrominance as suggested in [19].

Table 1 and Table 3 show periodicity computed for the walking and running sequences, respectively. Table 2 and Table 4 show a typical log-likelihood output for example walking and running sequences, respectively. Table 5 shows the recognition rate computed over all sequences. Figure 5 shows the typical segmentations from the walking sequences. Figure 6 shows typical segmentations from the running sequences.

Table 1 Walking periodicity.

	Computed Period	Actual Period
Person A Walk	16 frame	15 frame
Person B Walk	13 frame	13 frame
Person C Walk	15 frame	15 frame

We have found that due to the spatial median filtering in the motion segmentation step, the scale of the subject (which depends on a variety of factors including distance to the subject and focal length) has an influence on the

computed Hu moment feature values and hence on recognition accuracy. Because of the median filtering, some of the finer features of the subject silhouette are eliminated when the subject is at a smaller scale. Because our data for person C included several sequences taken at a greater distance than typical in our data, this effect is reflected in the lower recognition rate for person C .

Table 2 An example of the computed HMM log probabilities for three walk test sequences. Rows correspond to trained HMMs, and columns correspond to test sequences.

	Person A	Person B	Person C
Person A Walk	-74.9798	-470.282	-433.428
Person B Walk	-501.518	-107.909	-181.961
Person C Walk	-77.2505	-144.244	-124.003

Table 3 Running periodicity.

	Computed Period	Actual Period
Person A Run	9 frames	10 frames
Person B Run	12 frames	12 frames
Person C Run	10 frames	11 frames

Table 4 An example of the computed HMM log probabilities for three run test sequences. Rows correspond to trained HMMs, and columns correspond to test sequences.

	Person A	Person B	Person C
Person A Run	-55.1951	-61.4325	-132.271
Person B Run	-87.2595	-47.644	-118.383
Person C Run	-149.381	-112.923	-51.0995

Table 5 Recognition rate for individuals.

	Recognition Rate (%)
Person A	100% (16/16)
Person B	93.75% (15/16)
Person C	87.5% (14/16)

5. Conclusion

In this paper, a simple and reliable method is given for analyzing the periodicity of human activity and for recognizing individuals. Our approach is robust over variations in backgrounds, walking and running speeds, direction of motion (left or right), attire of the individuals, and, to some degree, distance to the camera. One limitation common to our approach and most previous approaches is that the motion must be frontoparallel. We expect that our approach would perform similarly in other fixed viewpoints if the training is also performed in the same viewpoint. An approach with a fixed viewpoint can find applications in surveillance (e.g., monitoring a hallway or an entrance where everyone passes in one of two orientations).

While the simplicity of this approach is appealing, it is unclear whether it will extend well to variable viewpoints or to larger numbers of individuals. One approach we will explore to expand the range of viewpoints is the use of affine invariant moments [18] rather than Hu moments. These features are invariant to more general transformations of the data, but the shorter feature vector may contain less information that is unique to individuals. The Hu moment features may also not be a rich enough feature set for distinguishing larger groups of individuals (although using time sequences certainly provides a high dimensional feature space). We will also explore richer feature sets such as joint angles, although these features may be difficult to extract reliably from video.

6. References

1. Yamato, J., J. Ohya, and K. Ishii. *Recognizing human action in time-sequential images using hidden Markov model.* in *Computer Vision and Pattern Recognition.* 1992. p. 379-385.
2. Brand, M., N. Oliver, and A. Pentland. *Coupled hidden Markov models for complex action recognition.* in *Computer Vision and Pattern Recognition.* 1997. San Jaun, PR.
3. Wilson, A.D. and A.F. Bobick, *Parametric Hidden Markov Models for Gesture Recognition.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999. **21**(9): p. 884-900.
4. Starner, T. and A. Pentland. *Visual Recognition of American Sign Language Using Hidden Markov Models.* in *International Conference on Automatic Face and Gesture Recognition.* 1995. Zurich, Switzerland.
5. Thrun, S., J.C. Langford, and D. Fox. *Monte Carlo hidden Markov models: Learning non-parametric models of partially observable stochastic processes.* in *ICML-99.* 1999.
6. Pentland, A. and A. Liu, *Modeling and Prediction of Human Behavior*, . 1995, M.I.T. Media Lab Perceptual Computing.
7. Bregler, C. *Learning and Recognizing Human Dynamics in Video Sequences.* in *Computer Vision and Pattern Recognition.* 1997. San Jaun, PR.
8. Wilson, A.D. and A.F. Bobick. *Nonlinear PHMMs for the Interpretation of Parameterized Gesture.* in *Computer Vision and Pattern Recognition.* 1998. Santa Barbara, CA.
9. Little, J.J. and J.E. Boyd, *Recognizing People by Their Gait: The Shape of Motion.* Videre: Journal of Computer Vision Research, 1998. **1**(2): p. 2-32.
10. Davis, J.W. and A.F. Bobick. *The Representation and Recognition of Action Using Temporal Templates.* in *Computer Vision and Pattern Recognition.* 1997. San Jaun, PR. p. 928-934.

11. Liu, F. and R.W. Picard. *Finding periodicity in space and time*. in *International Conference on Computer Vision*. 1998. Bombay, India.
12. Polana, R. and R. Nelson. *Detecting Activities*. in *Computer Vision and Pattern Recognition*. 1993. New York, NY. p. 2-7.
13. Tsai, P.-S. and M. Shah, *Cyclic Motion Detection*, . 1993, University of Central Florida.
14. Allmen, M.C. and C.R. Dyer. *Cyclic motion detection using spatiotemporal surfaces and curves*. in *International Conference on Pattern Recognition*. 1990. p. 365-370.
15. Gray, R.M., *Vector Quantization*. IEEE ASSP Magazine, 1984(4): p. 4-29.
16. Rabiner, L.R. and B.H. Juang, *An introduction to hidden Markov models*. IEEE ASSP Magazine, 1986(January 1986): p. 4-16.
17. Hu, M.K., *Visual pattern recognition by moment invariants*. IRE Transactions Information Theory, 1962. **8**(2): p. 179-187.
18. Sonka, M., V. Hlavac, and R. Boyle, *Image processing, analysis, and machine vision*. 2 ed. 1999: PWS Publishing.
19. Wren, C.R., *et al.*, *Pfinder: Real-time tracking of the human body*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997. **19**(7): p. 780-785.