RECOGNITION OF HUMAN AND ANIMAL MOVEMENT USING INFRARED VIDEO STREAMS

Qin Jiang and Cindy Daniell

Information Science Laboratory
HRL Laboratories, LLC
Malibu, California, USA

ABSTRACT

Distinguishing human motion from animal motion is important in many applications using infrared video streams, such as surveillance systems for homeland security and collision avoidance systems for nighttime driving safety. In this paper we present a technique to distinguish human motion from animal motion using infrared video sequences. In our technique, we uses frame differencing to represent object motion. Space-time correlation is used to characterize different type of motions. Our motion features are defined by Renyi entropy and mean values calculated from the correlations. A support vector machine-based classifier is used to classify the motion features. Our experimental results show that our technique is quite effective at distinguishing human motion from animal motion using infrared video sequences.

1. INTRODUCTION

With the advance of the infrared (IR) imaging sensor, IR sensors can now provide high quality images for both day and night applications. Because of this dual sensing capability, infrared video sequences are increasingly used in many applications, such as out-door surveillance systems and nighttime driving safety systems. Therefore, using infrared imagery to analyze / monitor human activity is becoming an important approach.

In the past decade, recognition of human activity has been a hot topic in computer vision research. Many techniques for recognizing human activity have been reported in the literature [1,2,3,4]. Unfortunately, most existing human activity recognition work has focused on using visual video sequences obtained from artificially well-controlled environments and lighting conditions. The techniques developed for visual video sequences are hardly applicable to infrared video sequences because of the differing image characteristics and qualities. Therefore, there is great demand for developing new techniques for recognizing human activities using IR video streams.

One of the important tasks for characterizing human activity is using IR video sequences to distinguish them from animal activities. In this paper, we present a technique to do this using only side-views of their motion sequences. In our technique, we first use image frame difference as a representation of motion information. Then, the motion information is characterized by the correlations computed in both the time and spatial domains. Two statistics, Renyi entropy and mean value, computed from the correlations are used as our motion features. Finally, a support vector machine-based classifier is used to classify the motion features. Recently, we collected real data to test our technique. The results show that our technique is very effective in distinguishing human motion from animal motion using infrared video sequences.

In the following section, we discuss the representation and extraction of motion information from infrared video imagery. Section 3 presents the characterization of the motion information using the space-time correlations and motion feature extraction. For classification of the motion features, a support vector machine-based classifier is given in section 4. Section 5 details our experimental tests and results. Finally, section 6 summarizes the paper and provides some conclusions as well as some future directions for research.

2. REPRESENTATION OF MOTION INFORMATION

Distinguishing human movement from animal movement can intuitively be regarded as distinguishing two-legged movement from four-legged movement. In this kind of approach, to represent the motion of moving legs, one needs to detect and track the moving legs in the image sequences. In the past decade, many techniques for limb tracking have been reported in the literature. However, to detect and track a moving object's limbs, most techniques require input images with a high quality, which can provide clear boundaries for the moving object shapes. Unfortunately, infrared video sequences of outdoor scenes are generally very noisy, especially when taken during the day. They hardly provide the image quality required for limb tracking. Therefore, using limb tracking to distinguish human and animal movements with infrared video sequences may not be a practical approach.

In the research on recognizing human activities using visual video sequences, motion fields are widely used to represent motion information about moving persons. However, one of the most significant differences between visual and infrared video images is that the latter provide no textual patterns and color information. It is well known that accurate estimation of motion fields will suffer if the images contain few textural patterns. Therefore, it is very difficult to compute accurate motion fields directly from infrared video sequences. One of the alternative ways is to use frame difference to extract motion information from video sequences. Although frame difference is generally less accurate than motion field data in terms of motion representation, it is computationally much simpler, which is very attractive in many real time applications. In addition, the effectiveness of motion representation has been

0-7803-8554-3/04/$20.00 ©2004 IEEE.

shown in many applications [4]. In our technique, we adopt frame difference to represent the motion information on both human and animal movements since accurate motion fields may not be computable from infrared video sequences as explained earlier. The computation of frame difference can be described mathematically as follows:

For a given infrared video sequence $f(x,y,t)$, frame difference is defined as

$$df(x,y,t) = |f(x,y,t+dt) - f(x,y,t)| \quad (1)$$

where x and y are spatial coordinates and t is the time index. The motion representation computed by frame difference is generally quite noisy. To remove the noise, we use a post-filtering process, which consists of thresholding and morphological filtering. Figure 1 shows several frames from a motion sequence and its motion representation.

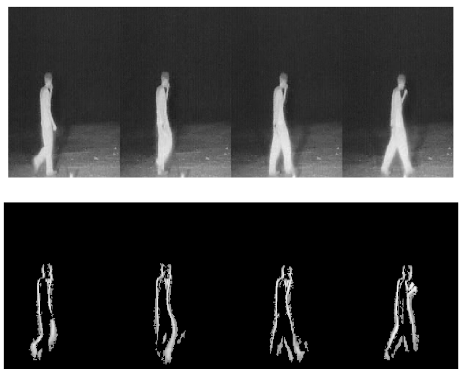

Figure 1: A motion sequence and its motion representation.

3. EXTRACTION OF MOTION FEATURES

Extracting motion features from the motion representation is the most important step in distinguishing human and animal movements. Whether the two kinds of movements can be well distinguished or not depends completely on whether motion features can characterize them effectively. In other words, good motion features can make the two kinds of movements well-separable in the feature space. As we mentioned before, one of the significant markers to separate human and animal movement involves the characteristics of two-legged vs. four-legged movement. Because detecting and tracking legs is not practical, we use correlations computed in a three-dimensional space, two spatial dimensions and one time dimension, to capture this distinction. Then, from the space-time correlations, we compute two statistics, Renyi entropy and mean value. We use these values as our motion features. The details are given in the following subsections.

3.1. Space-Time Correlations

With a sequence of frame differences as defined in Equation 1, we have a set of three-dimensional data, two dimensions in the spatial domain and one dimension in the time domain. We then compute space-time correlations along each dimension. Mathematically, the space-time correlations are defined as follows:

$$C_t(t) = Cor(df(x,y,t), df(x,y,t+\Delta t)) \quad (2)$$

$$C_x(x) = Cor(df(x,y,t), df(x+\Delta x,y,t)) \quad (3)$$

$$C_y(y) = Cor(df(x,y,t), df(x,y+\Delta y,t)) \quad (4)$$

where

$$Cor(u,v) = \frac{Cov(u,v)}{\sqrt{Var(u)Var(v)}} \quad (5)$$

The symbols $Cov(u,v)$ and $Var(u)$ are the covariance and variance of the variables. The space-time correlations provide a measure of the leg changes in the space-time domain. The two kinds of movements -- two-legged vs. four-legged -- are expected to show different patterns on their correlations.

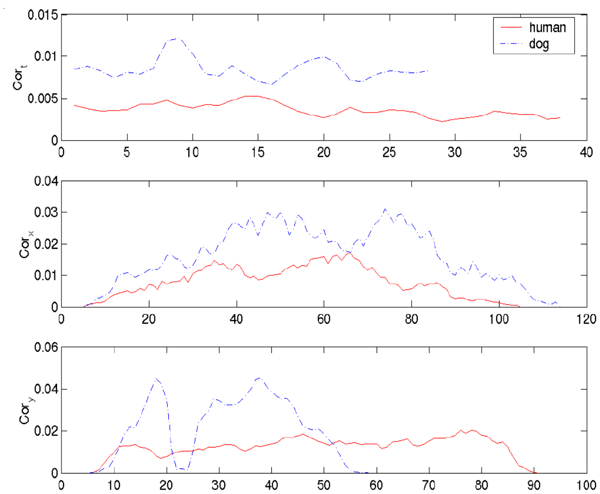

Figure 2: An example of space-time correlations of human vs. dog. Top: Cor-t, middle: Cor-x and bottom: Cor-y. Red lines are for human and blue dashed lines for dog.

Figure 2 presents the space-time correlations of a human motion sequence and a dog motion sequence. The correlations show some differences between the two sequences. For example, the correlations for the dog have larger variations than the ones for the human. The differences can be used as motion features to distinguish the two kinds of motion. Nevertheless, how to extract those differences is still an open problem. One possible approach is to use some statistics of the space-time correlations as motion features. In our technique, we use two statistics: mean value and Renyi entropy. They are described in the following subsection.

3.2. Motion Features

The space-time correlations present some different patterns between human and animal movements. We calculate two statistics, mean value and Renyi entropy, of the correlations to capture those different patterns. The mean value is simply defined by the following equation:

$$mean_s = E[Cor_s], \qquad s \in \{x, y, t\} \tag{6}$$

where $E[x]$ denotes the expectation of x.

The Renyi entropy is computed in the frequency domain. At first, we compute the Fourier transformation of the space-time correlations, that is

$$f_s = FFT(Cor_s), \qquad s \in \{x, y, t\}. \tag{7}$$

Then, we estimate the distributions $p_s(i)$ of f_s using a normalized histogram.

$$p_s = norm\_hist(f_s), \qquad s \in \{x, y, t\} \tag{8}$$

Finally, the Renyi entropy is calculated by the following equation:

$$h_s = \frac{1}{1-\alpha}\log_2\left(\sum_i (p_s(i))^\alpha\right), \qquad s \in \{x, y, t\} \tag{9}$$

where $\alpha > 0$.

The Renyi entropy, a generalized entropy, is a robust information measure [5]. When the order α approaches one, it equals the Shannon entropy. It has been used in many applications to provide a robust information measure [6].

4. CLASSIFICATION OF MOTION FEATURES

To classify the motion features, we use a support vector machine-based classifier. Here, we give a brief introduction about support vector machines (SVM). References [7,8] provide more information about SVM and its applications.

A support vector machine is a supervised statistical learning machine. In its learning process, an SVM constructs an optimal hyperplane as its decision surface using a small set of training data called support vectors, which are the data points closest to the decision surface and the most difficult to classify. The optimization for computing the decision surface is achieved in accordance with the principle of *structural risk minimization* [7]. Since in many applications, optimal decision surfaces could be non-linear, an SVM uses a set of non-linear transfer functions called the inner-product kernel to map the data from the input space into a high-dimensional feature space such that the non-linear decision surface in the input space becomes a linear decision surface (optimal hyperplane) in the feature space. Specifically, the procedure for constructing an SVM can be described as follows:

Let data set $\{(x_i, d_i)\}_{i=1}^{N}$ be the training data set and $K(x_i, x_j)$ be the inner-product kernel function. Then, the objective function of constructing an optimal decision surface is given as

$$J(\alpha) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j d_i d_j K(x_i, x_j) \tag{10}$$

subject to the constraints

$$\begin{aligned} &(1) \quad \sum_{i=1}^{N}\alpha_i d_i = 0 \qquad and \\ &(2) \quad 0 \le \alpha_i \le C \qquad for \qquad i = 1,2,3.....,N \end{aligned} \tag{11}$$

where C is a user-specified positive parameter called the "cost" of mistakes. The optimal vector α^* is determined by maximizing $J(\alpha)$. Then, the equation for the optimal decision surface is written as

$$f(x) = \sum_{i=1}^{N_s}\alpha_i^* d_i K(v_i, x) + b \tag{12}$$

where v_i is a support vector, N_s is the number of support vectors, and b is the bias term. In our techniques, we use the Gaussian function as the kernel function, given by

$$K(x, y) = \exp\left(-\frac{1}{2\sigma^2}\|x - y\|^2\right) \tag{13}$$

We chose this approach because compared to other classification techniques such as neural networks and decision trees, SVM can provide a better generalization performance on pattern classification problems [8].

5. EXPERIMENTAL TEST

We used 38 human motion sequences and 38 animal motion sequences, 12 dog sequences and 26 horse sequences, to test our technique. The lengths of the sequences vary from 25 frames to 45 frames depending on the speed of the moving objects (human or animal) in the sequences. A Raytheon infrared camera Palm-IR 250 (7-14 micron uncooled ferro-electric type) was used to collect the testing sequences. Each image frame has a resolution of 320 x 240 pixels, and each sequence has a sampling rate of 30 frames per second. Some sequences were taken at night, and some were taken in the daytime. All of them were obtained from outdoor scenes. As a result, the backgrounds in the sequences are very noisy and the sizes of the moving objects are quite different. Figure 5 presents some samples of the testing data. From the data samples, we can see that in most cases, it is almost impossible to detect the number of legs of the moving targets. We used a target window to extract moving objects. The size of the window depended on the size of the objects. The space-time correlations were computed using the windowed 3D object data, two dimensions in the spatial domain and one dimension in the time domain. The values of the space-time correlation are not sensitive to the size of the moving objects. To calculate the feature for Renyi entropy, we used parameter $\alpha = 0.1$ and 1024 point FFT. From the clusters of the features, we can see that the two types of motions are well separable in the feature space, especially in the subspace of the feature of Renyi entropy. Figure 3 and Figure 4 show the cluster plots of our features. A support-vector-machine-based classifier was used for motion feature classification. For each class, 19 sequences were used to train the classifier and 19 sequences were used for testing. The results of the classification are summarized in Table 1. On average, our system can achieve a classification rate of 92.5%, which is quite impressive considering the quality of the testing data.

Table 1: Confusion Matrix of Feature Classification

	Computed Classes	
True Classes	Human Motion	Animal Motion
Human Motion	0.95	0.05
Animal Motion	0.10	0.90

6. CONCLUSIONS

In this paper, we developed a human motion recognition system using IR video streams. Our system uses image frame differencing to represent object motion. The changes of the different motions of human vs. animal are captured by space-time correlations. With the space-time correlations, two statistics, Renyi entropy and mean value, are calculated as motion features. A support vector machine-based classifier is used to classify the motion features. Our experimental tests show that with our motion features, our human motion recognition system can achieve a very good classification rate, namely, an average classification rate of 92.5% even though the testing data is very noisy.

7. REFERENCES

[1] J. W. Davis and A. F. Bobick, "The representation and recognition of human movement using temporal templates," in *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 928-934, 1997.

[2] James J. Little and Jeffrey E. Boyd, "Recognizing people by their gait: the shape of motion," *Videre: Journal of Computer Vision Research*, vol. 1, No. 2, pp. 1-32, 1998.

[3] J. Yamato, J. Ohya and K. Ishii, "recognizing human action in time sequential images using Hidden Markov Model," in *Proc. of IEEE Conf. On CVPR,* pp. 379-385, 1992.

[4] Osama Masoud and Nikos Papanikolopoulos, "Recognition human activities," in *Proc. of IEEE Conf. on CVPR*, pp. 157-162, 2003.

[5] A. Renyi, "*Probability Theory*," Amsterdam the Netherlands: Northholland, 1970.

[6] R. G. Braraniuk, P. Flandrin, A. J.E.M. Janssen and O. J.J. Michel, "Measuring time-frequency information content using the Renyi entropies," *IEEE Trans. on Information Theory*, vol. 47, No. 4, pp. 1391-1409, 2001.

[7] V. Vapnik, "*The Nature of Statistical Learning Theory*," Spring-Verlag, New York, 1995.

[8] V. Vapnik, S. Golowich, and A. Smola, "Support vector method for function approximation, regression estimation, and signal processing," *Advance in Neural Information Processing Systems*, vol. 9, pp. 281-287, 1996.

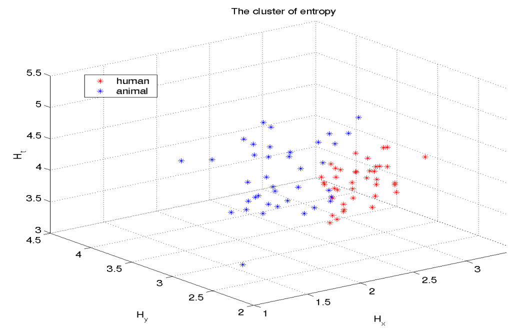

Figure 3: Cluster of Entropy
Red for human and blue for animal

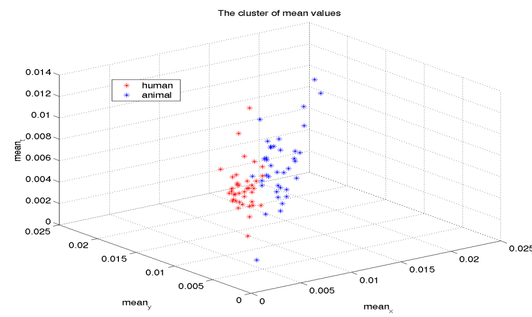

Figure 4: Cluster of Mean Value
Red for human and blue for animal

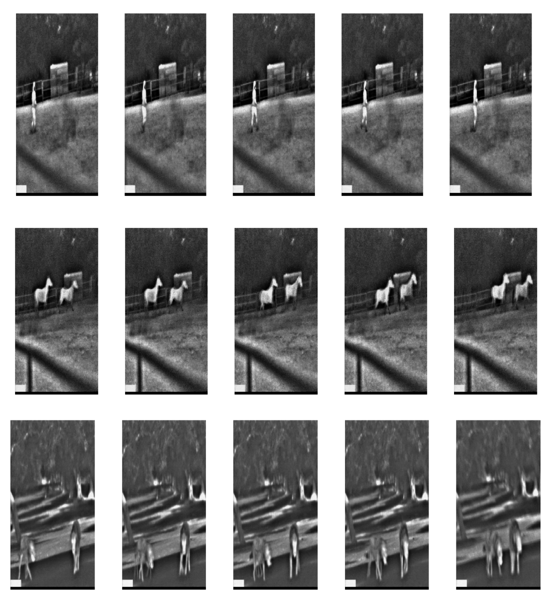

Figure 5: Test Data
Top: human, middle: horse and bottom: dog.