Human Activity Recognition Using Multidimensional Indexing

Jezekiel Ben-Arie, *Member*, *IEEE*, Zhiqian Wang, *Member*, *IEEE*,
Purvin Pandit, *Member*, *IEEE*, and Shyamsundar Rajaram, *Student Member*, *IEEE*

Abstract—In this paper, we develop a novel method for view-based recognition of human action/activity from videos. By observing just a few frames, we can identify the activity that takes place in a video sequence. The basic idea of our method is that activities can be positively identified from a sparsely sampled sequence of a few body poses acquired from videos. In our approach, an activity is represented by a set of pose and velocity vectors for the major body parts (hands, legs, and torso) and stored in a set of multidimensional hash tables. We develop a theoretical foundation that shows that robust recognition of a sequence of body pose vectors can be achieved by a method of indexing and sequencing and it requires only a few pose vectors (i.e., sampled body poses in video frames). We find that the probability of false alarm drops exponentially with the increased number of sampled body poses. So, matching only a few body poses guarantees high probability for correct recognition. Our approach is parallel, i.e., all possible model activities are examined at one indexing operation since all of the model activities are stored in the same set of hash tables. In addition, our method is robust to partial occlusion since each body part is indexed separately. We use a sequence-based voting approach to recognize the activity invariant to the activity speed. Experiments performed with videos having eight different activities show robust recognition with our method. The method is also robust in conditions of varying view angle in the range of $\pm$ 30 degrees.

Index Terms—Human activity recognition, multidimensional indexing, sequence recognition, human body part tracking, EXpansion Matching (EXM).

1 Introduction

Human activity recognition from video streams has a wide range of applications such as human-machine interaction, choreography, sports, security surveillance, content-based retrieval, etc. Depending on the environment, human activity may have different forms ranging from simple hand gestures to complex dances. In this paper, we focus on human activity recognition based on angular poses and velocities of the main human body parts. Even though we briefly explain the method for body part tracking and other related issues, the main contribution of this paper is a novel recognition method of human activity that is more efficient than prevailing methods.

The human body's pose frequently gives an indication of the action that takes place. In Fig. 1, we present an example of a simple activity. It is not hard to determine that the activity depicted is of walking. One can classify the activity just by looking at Fig. 1. This classification does not require a full video sequence and only three samples are sufficient to classify the activity with high certainty. This is the idea that led us to develop our indexing-based method.

In our work, we are focusing on gross activities (such as walking, jumping) that entail motion of major body parts, i.e., the arms, legs, torso, and head. Human activity can be described as a temporal sequence of pose vectors that represent sampled poses of these body parts. Our principle of recognizing human activity from sparsely sampled poses is based on identifying these poses as samples of a complete, densely sampled model activity. To achieve this objective, we construct a database for all the major body parts that includes all the model activities in the form of pose entries in multidimensional hash tables. Each body part has a separate hash table which includes all the model activities. The poses of the body parts are represented by a set of normalized body part angles to achieve invariance to body size. Hence, models of human activity are represented by a sequential arrangement of sets of multidimensional vectors that correspond to sampled angular poses[1] of body parts over the entire time interval. These vectors are then divided into a set of subvectors where each subvector corresponds to the angular pose of different body part. Next, we form a set of hash tables, each of which corresponds to an individual body part. The indices in these hash tables are the poses of the corresponding body parts (the subvectors) and the contents of these hash tables are the identities of the model activities and their time labels. The size of these tables is not too large since body parts have limited angular motion and, thus, the number of bins that describe the full range of angular motion of each body part is quite limited. An important feature of our approach is the separation of the multidimensional indexing into several hash tables, where each table corresponds to a different body part. This structure enables us to index and recognize activities even when several body parts are occluded (as elaborated in Section 5). Also, our approach of using multidimensional vectors proves to be very efficient

• *The authors are with the ECE Department M/C 154, University of Illinois at Chicago, 851 S. Morgan St., Chicago, IL 60607.*
E-mail: {benarie, srajaram}@ece.uic.edu. zwang@RCTanalytics.com. purvin@ieee.org.

Manuscript received 30 Jan. 2001; revised 31 July 2001; accepted 17 Jan. 2002.
Recommended for acceptance by S. Sarkar.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 113551.

1. By the terms "angular poses" and "body poses," we refer to angular poses and angular velocities of the nine major body parts.

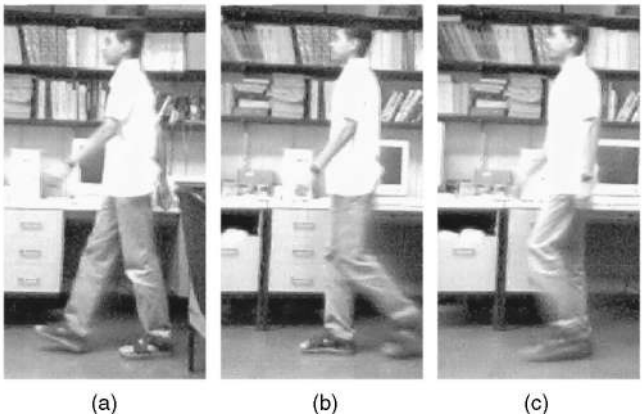


Fig. 1. Three sampled video frames of a man walking. (a) Frame 7, (b) frame 12, and (c) frame 15.

in terms of storage since all the activities are stored in the same set of hash tables.

At the recognition stage, a set of multidimensional indices (which correspond to angular poses) are derived from the video sequence of the test activity. Each video frame in the test activity sequence yields a vote vector for each activity model in the database. This vote vector is a temporal depiction of the log-likelihood that the indexed pose belongs to the activity model. The overall vote for each activity model is obtained by integrating the votes for all the test frames using sequential correlation. Details are provided in Section 4. The overall votes for each activity model correspond to the likelihood that the test frames actually are samples of this particular model activity. As elaborated in Section 3, our approach is robust to variations in an activity's speed and timing. Another advantage of this approach is the tremendous flexibility it provides in sampling activity sequences. There are no strict requirements either on the number of sampled body poses (frames) or on the frame intervals of the test sequence. Users need only a set of sparsely sampled representative body poses for activity recognition. We show in Section 3 that the probability for false detection declines exponentially with the number of sampled body poses. So, only a few body poses are required. The organization of the activity database also results in tremendous reduction of recognition time and of space requirements since all the candidate model activities are examined at one voting action as they are stored in the same set of hash tables.

In Section 3, we describe the theory behind our approach. In Section 4, we describe the multidimensional indexing and the voting schemes. Section 5 discusses the application of our approach and demonstrates the method experimentally. The main contribution in this paper is a novel recognition method and our tracking scheme only provides inputs; therefore, it is elaborated in the appendix.

2 Previous Work

A paper by Gavrila [10] is an excellent reference which conducts an intensive survey on the different methodologies for visual analysis of human movement. Gavrila groups them into 2D approaches with or without explicit shape models and 3D approaches. The 2D approach without explicit shape models is based on describing human movement in terms of simple low-level 2D features instead of recovering the pose. The second approach, which is a view-based approach, uses explicit shape models to segment, track, and label body parts. The third approach attempts to recover the 3D poses over time. More recently, there has been a survey by Moeslund and Granum [17] which describes various computer vision-based human motion capture. They elaborate about the various categories of human motion capture, namely, initialization, tracking, pose estimation, and recognition. Human motion recognition is classified into static and dynamic recognition. Static recognition is based on using spatial data, one frame at a time, and dynamic recognition uses the temporal characteristics of the action. Our indexing and sequencing-based approach differs from all the methods surveyed by Gavrila [10] and Moeslund and Granum [17] since it combines the static and dynamic recognition approaches.

Unlike our work, which can classify eight different activities and can be easily extended to even more than eight activities, past works focused on recognition of very few activity classes. Fujiyoshi and Lipton [8] use skeletonization to extract internal human motion features and to classify human motion into "running" or "walking" based on the frequency analysis of the motion features. Yang and Ahuja [28] apply time-delay neural network (TDNN) to hand gesture recognition and achieve quite a high recognition rate. Schlenzig et al. [24] use Hidden Markov Model (HMM) and a rotation-invariant imaging representation to recognize visual gestures such as "hello" and "good-bye." HMMs are used by Yamato et al. [27] for recognizing human action in time sequential images. HMMs are also utilized by Starner and Pentland to recognize American Sign Languages (ASL). Darrell and Pentland apply dynamic time warping to model correlation for recognizing hand gestures from video. Polana and Nelson [22] use template matching techniques to recognize human activity. Motion Energy Images are used by Bobick and Davis [1] for recognition.

Haritaoglu et al. [11] implemented a system for human tracking and activity recognition in which the activity recognition part is mainly based on analysis of the projected histograms of detected human silhouettes. This system classifies human poses in each frame into one of four main poses (standing, sitting, crawling/bending, lying) and one of three view-based appearances (front/back, left side, and right side) and activities are monitored by checking the pose changes over time. In another work, Ivanov and Bobick [12] recognize generic activities using HMM and stochastic parsing. These activities are first detected as a stream of low level action primitives represented using HMM and then are recognized by parsing the stream of primitive representations using a context-free grammar. Bobick and Davis [2] recognized human activity by matching temporal templates against stored instances of views of known actions. More recently, Galata et al. [9] use Variable-Length Markov Models (VLMM) for modeling human behavior. They use VLMMs because of their more powerful encoding of temporal dependencies. Our indexing-based recognition approach differs from all the above-mentioned works since it determines the best matching activity in a single indexing operation.

3 Theoretical Foundation of Our Approach

In this section, we describe the theoretical foundation to our approach in recognizing human activity using indexing. Our representation for human activity in video frames could be described as a concatenation of 18-dimensional subvectors $\mathbf{x_i}$ that describe the angles and angular velocities of nine body parts.[2] Each subvector pertains to a video frame and, thus, the whole video sequence can be represented by a vector $\mathbf{Y}$ which is a concatenation of all the subvectors $\mathbf{x_i}$. Please note that, in our representation, the angles are only 2D projections of the actual 3D angles. Hence, our representation is limited to a given view of the activity and, so, our scheme is view-based. However, we find that this representation is not very sensitive to changes in vantage point and the viewing direction can be changed in the range of ± 30 degrees without seriously affecting the recognition rate. Experimental results that verify this assertion are included in Section 5 and in Fig. 17. In the future, we plan to incorporate a method for recovery of the 3D angles [7] that will enable us to make our recognition method view invariant.

To recognize an activity, one has to compare the test video to a model activity. In other words, the test vector $\mathbf{Y_t}$ has to be compared with a set of model vectors $\{\mathbf{Y_m}; m \in [1, M]\}$, where M is the number of activity models in the database. A similar problem was dealt with using Hidden Markov Models (HMM) [12], [19], [24]. We find that the solution can be significantly simplified if we make some assumptions that will be detailed later. The problem of activity recognition can be formulated as a Maximum Likelihood Sequence Estimation (MLSE). The MLSE problem is to determine the most likely sequence $\mathbf{Y_m}$ given the observations $\mathbf{Y_t}$. The Viterbi algorithm [18] provides a computational approach to solving such a problem. However, we use an indexing approach which is computationally simpler. We assume that the random differences between the subvectors $\mathbf{x_t}$ and $\mathbf{x_m}$ can be described as multivariate zero mean Gaussian distribution. Assuming that these variations are conditionally independent from sample to sample, then the likelihood function for the sequence $P(\mathbf{Y_t}|\mathbf{Y_m})$ can be written as

$$P(\mathbf{Y_t}|\mathbf{Y_m}) = P(\mathbf{x_{t_1}}, \mathbf{x_{t_2}}, \cdots, \mathbf{x_{t_k}}|\mathbf{x_{m_1}}, \mathbf{x_{m_2}}, \cdots, \mathbf{x_{m_k}}) = \prod_{i=1}^{k} \frac{e^{\left[\frac{-1}{2}(\mathbf{x_{t_i}}-\mathbf{x_{m_i}})^T C_x^{-1}(\mathbf{x_{t_i}}-\mathbf{x_{m_i}})\right]}}{(2\pi)^{\frac{N}{2}}|C_x|^{\frac{1}{2}}}, \quad (1)$$

where C_x is the covariance matrix of the distribution of the training set for $\mathbf{x_m}$, N is the dimension of the subvectors $\mathbf{x_m}$ or $\mathbf{x_t}$ (18 in our case), and k is the number of frames in the activity sequence. Using the log-likelihood function, we get

$$\log P(\mathbf{Y_t}|\mathbf{Y_m}) = \sum_{i=1}^{k}\left[\frac{-1}{2}(\mathbf{x_{t_i}}-\mathbf{x_{m_i}})^T C_x^{-1}(\mathbf{x_{t_i}}-\mathbf{x_{m_i}})\right] - kG, \quad (2)$$

where G is the logarithm of the denominator in (1) given by

$$G = \log\left[(2\pi)^{\frac{N}{2}}|C_x|^{\frac{1}{2}}\right]. \quad (3)$$

2. The nine body parts are torso and head, upper arms and legs (thighs) and lower arms (forearm plus hand), and legs (calf plus foot).

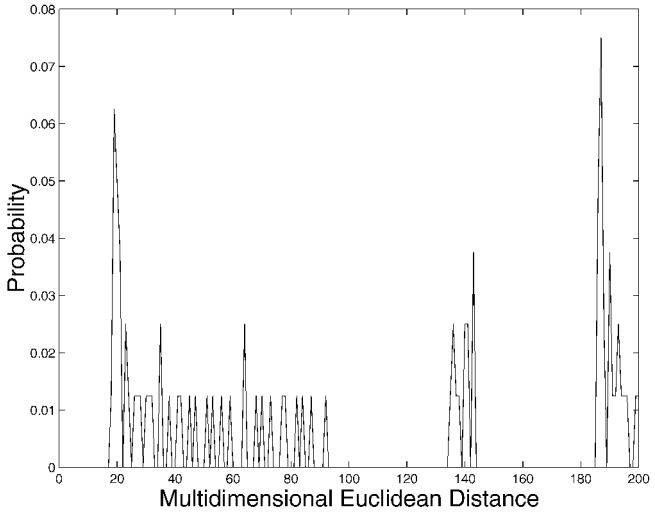


Fig. 2. Probability distribution of the multidimensional Euclidean distance for the jumping activity and the sitting activity. This distribution is quite sparse compared to the one in Fig. 3.

The most likely activity sequence Ω is found by the maximum-likelihood approach,

$$\Omega = \arg\max_m \left(\sum_{i=1}^{k}\left[\frac{-1}{2}(\mathbf{x_{t_i}}-\mathbf{x_{m_i}})^T C_x^{-1}(\mathbf{x_{t_i}}-\mathbf{x_{m_i}})\right]\right). \quad (4)$$

3.1 Foundations of the Voting Approach

Finding the most likely activity can now be solved by an indexing-based voting approach. In this case, for each test subvector $\mathbf{x_{ti}}$, we accumulate votes for all the models. In such voting, a model m will accumulate an incremental vote of

$$\frac{-1}{2}(\mathbf{x_{t_i}}-\mathbf{x_{m_i}})^T C_x^{-1}(\mathbf{x_{t_i}}-\mathbf{x_{m_i}}) - G \quad (5)$$

for each test frame i. This process is repeated by voting for all the frames i in the test sequence. In our method, we even simplify this voting further by **voting only on a few representative frames** which are sparsely sampled from the test video sequence.

In order to assess how many sampled poses/frames are necessary for robust activity recognition, in general, one has to characterize the behavior of the probability of false matching as a function of the number of sampled test body poses K. We first perform a statistical study of the probability distribution $p_{AB}(\delta)$ of the minimal multidimensional distance δ between activities A and B. This is done simply by finding, for every frame of A, the minimal distance to any frame in B. In Fig. 2, we show $p_{AB}(\delta)$, where A is the jumping activity and B is the sitting activity. In Fig. 3, we show $p_{AB}(\delta)$, where both A and B are jumping activities but performed by different people. This probability distribution curve is a measure of the closeness between the two activities when the number of sampled test poses is one. The high density of the probability distribution in the range of lower values of distances for the two different jumping sequences in Fig. 3 clearly show the closeness between the two versions of the jumping activity even when they are performed by different people, whereas the probability distribution of the jumping and sitting activity in Fig. 2 is comparatively sparse. The area under this probability distribution curve $p_{AB}(\delta)$ below a threshold γ,

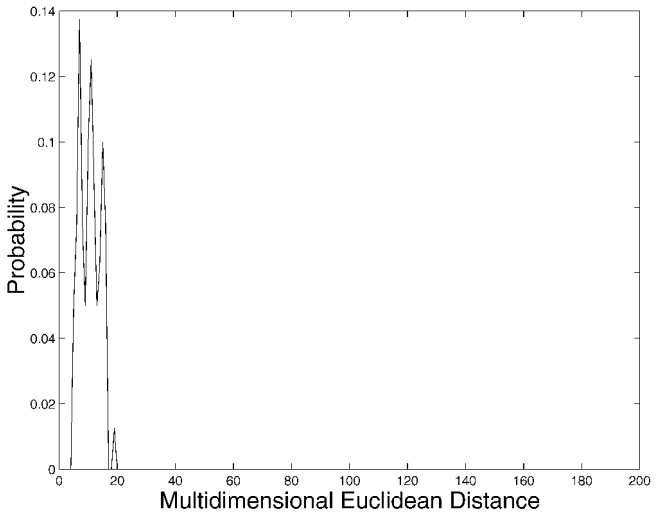


Fig. 3. Probability distribution of the multidimensional Euclidean distance between two different versions of jumping performed by different persons. Note that this distribution is quite concentrated in the low distance values compared with the one in Fig. 2.

$P_{AB}(\delta \leq \gamma)$, indicates the probability that the test sequence is identified incorrectly (assuming that A $\neq$ B). This threshold γ is the distance that corresponds to a good match.

$$P_{AB}(\delta \leq \gamma) = \int_0^{\gamma} p_{AB}(\delta)\, d\delta = \alpha. \tag{6}$$

Assuming that P_{AB} is independent from sample to sample, the joint probability for a false matching in all the frames when using K test sampled poses can be obtained by,

$$P_{AB}(\delta_1 \leq \gamma \; and \; \delta_2 \leq \gamma \cdots \delta_K \leq \gamma) \leq \alpha^K. \tag{7}$$

The inequality arises from our sequencing method that reduces the probability for false matching by further reducing the time interval for each subsequent sample. For example, we show in Fig. 4 the probability for false match between jumping and kneeling as a function of K. Actually, the sequencing process further reduces the probability for false matching since each subsequent frame is matched in a decreased space and, therefore, has even lower probability $P_{AB}(\delta)$.

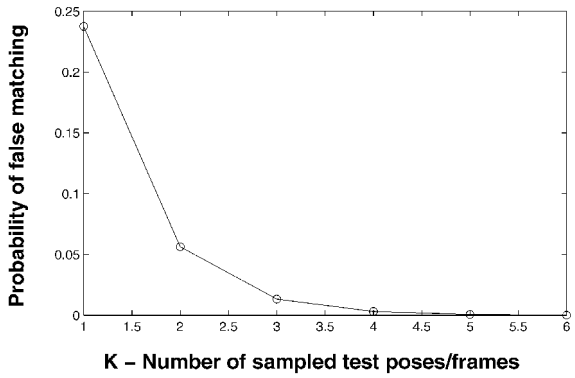


Fig. 4. Probability for false matching between jumping and kneeling as a function of K.

3.2 Dealing with Time Shifts and Activity Speed Variations

In most test sequences, we encounter the problem that the activity is not synchronized with the model activity. Usually, there is a time shift between the two sequences. This time shift denoted by a, is a priori unknown and has to be found, along with the activity classification. We solve this problem by combining the votes with temporal correlation.

$$\Omega = \arg\max_m \left(\arg\max_{a_m} \left(\sum_{i=1}^{k} \left[\frac{-1}{2} \left(\mathbf{x}_{\mathbf{t_i}} - \mathbf{x}_{\mathbf{m_{i-a_m}}} \right)^T C_x^{-1} \left(\mathbf{x}_{\mathbf{t_i}} - \mathbf{x}_{\mathbf{m_{i-a_m}}} \right) \right] \right) \right), \tag{8}$$

where a_m is the time shift between the test sequence and the mth model sequence of the activity.

Another problem that arises in many activities is the problem of speed variations of the activity. The same activity could be performed with different speeds and the speed can even vary during the course of the activity. Variations of speed are actually equivalent to variations in time scale. This problem is quite difficult, in general, since it requires complex search for the optimum votes with various time scales and time shifts.

$$\Omega = \arg\max_m \left(arg\max_s \left(\arg\max_{a_m} \left(\sum_{i=1}^{k} \left[\frac{-1}{2} \left(\mathbf{x}_{\mathbf{t_i}} - \mathbf{x}_{\mathbf{m_{s(i-a_m)}}} \right)^T C_x^{-1} \left(\mathbf{x}_{\mathbf{t_i}} - \mathbf{x}_{\mathbf{m_{s(i-a_m)}}} \right) \right] \right) \right) \right), \tag{9}$$

where s denotes the time scale.

In Section 4, we propose a method which provides an efficient and robust solution to speed invariant activity recognition. Our solution is based on sequence matching of the sparse samples. The first underlying principle in the method is that the sequence of the samples of any activity do not change with any variations of speed. This is obvious. Thus, we can reduce the search space by first searching for the optimal vote for the first test frame $\mathbf{x}_{\mathbf{t}_1}$ and then searching for the next optimal vote for the second test frame $\mathbf{x}_{\mathbf{t}_2}$ only in the reduced set of model frames which occur **after** the matched model frame with $\mathbf{x}_{\mathbf{t}_1}$. The same process repeats with the third test frame, the fourth test frame, and so on.

In periodic activities (such as walking, running, etc.), the first frame in the test activity may match a model frame which is toward the end of the period and the second test frame will then match a model frame which occurs before the first matched model frame. Imposing the sequencing condition in such a case will cancel such matching. To avoid such situations, we extend each model activity sequence to two consecutive periods for all the periodic activities.

4 Multidimensional Indexing and Voting

In this work, we are interested in activities that involve motions of major body parts. Therefore, the human body is represented only by nine generalized cylinders for the torso, upper arms and legs (thighs), and lower arms (forearms + hands) and legs (calves + feet), as in Fig. 5. In our video analysis, we consider the 2D projections of this model. The 2D Cartesian coordinates of all the major joints connecting the

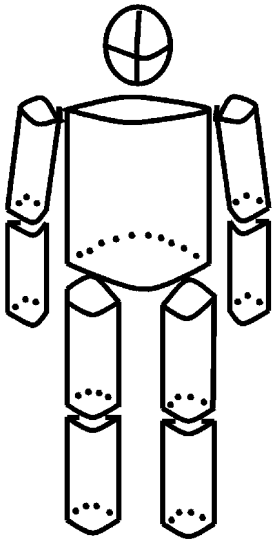


Fig. 5. A 3D human body model.

above mentioned parts are derived using a tracking procedure for body parts which is explained in the Appendix. For our application, the pose of the whole body at any instant is composed of the poses of the arms, legs, and torso. To achieve invariance to body size, the 2D Cartesian coordinates are transformed into 2D angles. The hash table is four-dimensional for the limbs and two-dimensional for the torso. We use 4D tuples $(\theta_1, \theta_2, \dot{\theta}_1, \dot{\theta}_2)$, where θ_1 denotes the angle between the positive x-axis and the upper arm or the thigh and θ_2 represents the angle between the positive x-axis and the forearm or the calf, $\dot{\theta}_1$ denotes the angular velocity of the upper arm or thigh and $\dot{\theta}_2$ denotes the angular velocity of the forearm or the calf. For the torso, the 2D tuples are $(\theta_3, \dot{\theta}_3)$, where θ_3 represents the angle between the positive x-axis and the major axis of the torso and $\dot{\theta}_3$ is the angular velocity of the torso. The angular velocities are calculated as the difference of the angular positions of two successive frames.

The next step is to quantize these multidimensional tuples into multidimensional bins to form indices into separate hash tables. In our indexing scheme, we have five hash tables: one (h_1) for the torso, two (h_2 and h_3) for the legs, and two (h_4 and h_5) for the arms. The model information stored in the hash table contains a pair of values which denote the model number $\{m; m \in [1, M]\}$ and the time instant $\{t; t \in [0, T_m - 1]\}$ of the model activity in the database, where M is the number of activity models in the database and T_m represents the number of image frames for model m. This information is stored at the bins that correspond to the angular pose that pertains to the model activity m at the particular instant t. The hash table structure for the limbs is shown in Fig. 6. Each hash table is updated using the angular position of the body parts obtained from each activity model. In the hash table, every entry may include a set of different activity models which pertain to the same body part pose. This arrangement of the hash tables is quite efficient for storage since it includes all the model activities in the same table and also enables robust recognition.

Our recognition scheme consists of three stages: The first stage involves voting for the individual body parts. The second stage combines the votes of the individual body parts for each test pose/frame. The third stage obtains the final activity vote by integrating the votes of individual test frames based on the sequence information. The recognition scheme is illustrated in Fig. 7 in the form of a flow diagram.

In the first stage, we decompose the body pose in each frame into angular poses and velocities of body parts and index into the hash tables of the corresponding parts. The voting scheme for each part h_i employs M 1D arrays $V_{mk}^{h_i}(t), m \in [1, M]$, where each array corresponds to a different activity model and to k, which is the frame number of the test activity. One may have several items in the same hash table bin that correspond to the same pose index; such items may correspond to different activity models and/or may pertain to different time instants. In order to tolerate slight pose variations that may occur in the same activity, it is necessary to also consider the neighboring pose bins of the indices derived from the poses of the test activity. Let $b_i^k = (q_1^k, q_2^k, q_3^k, q_4^k)$ denote the quantized bin of one of the limbs $(h_i, i \in [2, 5])$ for a test pose in test frame k and let $b_i' = (q_1', q_2', q_3', q_4')$ denote a neighboring bin in the corresponding hash table. We define $f(b, c, d, e)$ as a mapping function from a bin's offset b, c, d, e to the f range $[0, -\infty)$. Here, we choose

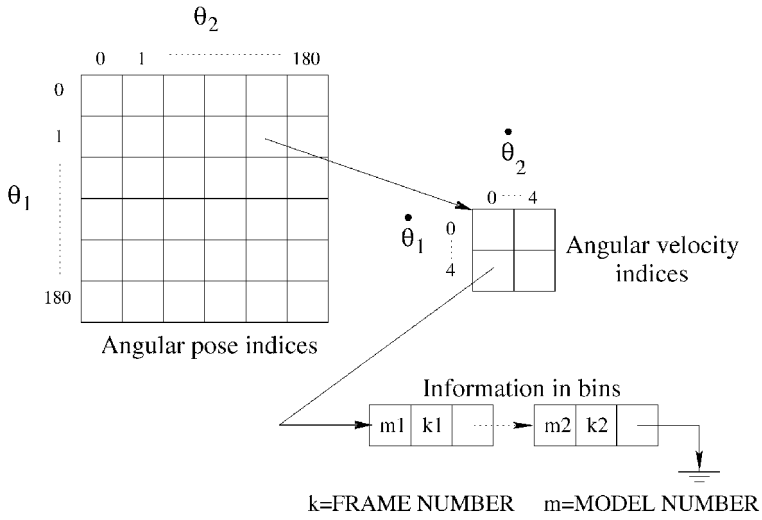


Fig. 6. The hash table structure used for the limbs.

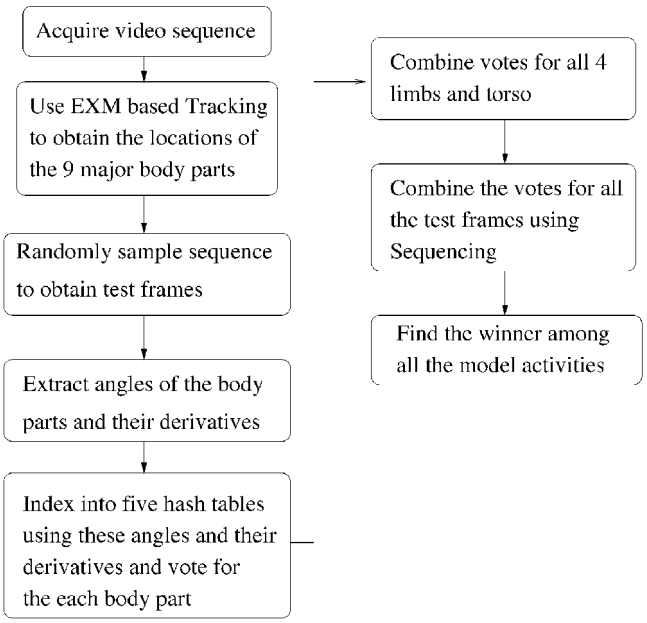


Fig. 7. The Flow Diagram of our recognition approach.

this mapping function to be a logarithm of a 4D Gaussian (assuming uncorrelated covariance matrix C_x) which conforms with our assumed model in (2),

$$f(b,c,d,e) = log\left[e^{\frac{-1}{2}\left[\left(\frac{b-b_0}{\sigma_b}\right)^2+\left(\frac{c-c_0}{\sigma_c}\right)^2+\left(\frac{d-d_0}{\sigma_d}\right)^2+\left(\frac{e-e_0}{\sigma_e}\right)^2\right]}\right], \quad (10)$$

where $\sigma_b, \sigma_c, \sigma_d, \sigma_e$ denote the scale of the Gaussian along the respective axes, (b_0, c_0, d_0, e_0) represent the center of the function. The standard deviations of the Gaussians are selected using statistical analysis. Each model activity is derived by averaging several video sequences containing the same activity. Since the number of frames in each of these sequences is different in most instances, we interpolate between frames to make this number equal. We then calculate the variance of the angular pose for each body part over all the sequences for that activity at that frame instant. This is then done for all the frames of the sequences taken for that activity and all these variances are averaged to get a single variance for each body part for that activity. These variances are later used in our probability distribution models (as in (10)). We do this for all the models and for all the body parts.

In the voting process, a model m with time instant t in the entry $h_i(b_i'), i \in [2,5]$ receives a vote from the test pose b_i^k of one of the limbs according to

$$V_{mk}^{h_i}(t) \mathrel{+}= \big(f\big(|q_1^k - q_1'|, |q_2^k - q_2'|, |q_3^k - q_3'|, |q_4^k - q_4'|\big)\big), \quad (11)$$

where $+=$ represents incrementing the value of the left-hand side by the value of the right-hand side. $V_{mk}^{h_i}(t)$ is initialized to zero before the voting begins. f is defined in (10) and q_1^k, q_2^k, q_3^k, and q_4^k and q_1', q_2', q_3', and q_4' are defined in a prior paragraph. This voting mechanism is illustrated in Fig. 8. For additional voting on the poses of the torso, we use

$$V_{mk}^{h_1}(t) \mathrel{+}= f\big(|q_5^k - q_5'|, |q_6^k - q_6'|\big), \quad (12)$$

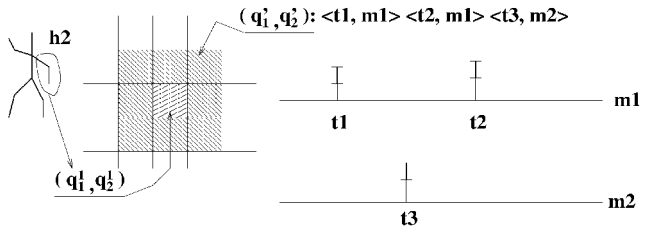


Fig. 8. A voting example of the left arm. On the left, the center square (q_1^l, q_2^l) of the grid represents the bin from the pose of the left arm and the surrounding squares are neighboring bins. The upper-right bin (q_1', q_2') contains three entries from models $m1$ and $m2$. These votes are described by the bars on the right diagram. This diagram describes two 1D voting arrays for activity models $m1$ and $m2$.

where q_5^k denotes the quantized bin of the angular pose of the torso in test frame k, q_6^k denotes the quantized bin of the angular velocity of the torso in test frame k, q_5', q_6' denote a neighboring bin, and the mapping function f is a logarithm of a 2D Gaussian in this case.

In the second stage, the votes that correspond to a particular test image frame k are denoted as $V_{mk}(t)$ and are obtained by combining the votes for the torso and the votes for other body parts. The votes from the limbs and torso are combined by addition. Hence, the votes for a test image frame are given by:

$$V_{mk}(t) = \sum_{i=1}^{5} V_{mk}^{h_i}(t). \quad (13)$$

This process of voting and combining of votes for different body parts is illustrated in Fig. 9. The final result of the first two stages is a set of M 1D voting arrays $\{V_{mk}(t); m = 1, \cdots, M\}$, where m is the model number and k represents a test frame $k = 1, \cdots, K$ and K is the number of test frames. A numerical example of the first two stages of voting is shown in Fig. 10.

In the third stage, the votes obtained from all the individual test frames are combined. This combination can be done in two ways, temporal or sequential correlation. In temporal correlation, the temporal difference between the successive test frames is used to combine the votes. This method fails for activities performed at different speeds. For speed invariant activity recognition, we find that, even though the speed of the activity may change, the sequence of the body poses always remains almost the same (for the same activity). So, we can combine the votes of the test frames by using sequence information as follows: The final vote for the mth model can be obtained by the following equation

$$V_m = \sum_{k=1}^{K} V_{mk}(L_k). \quad (14)$$

In order to have the same sequences, the following conditions have to be satisfied in (14): $L_i < L_j$; $i < j$ and $V_{mk}(L_k)$ is the maximum vote for the activity m and L_k is the argument of the maximum vote. The recognized activity is then the model that corresponds to the maximum of $V_m; m = 1, \cdots, M$.

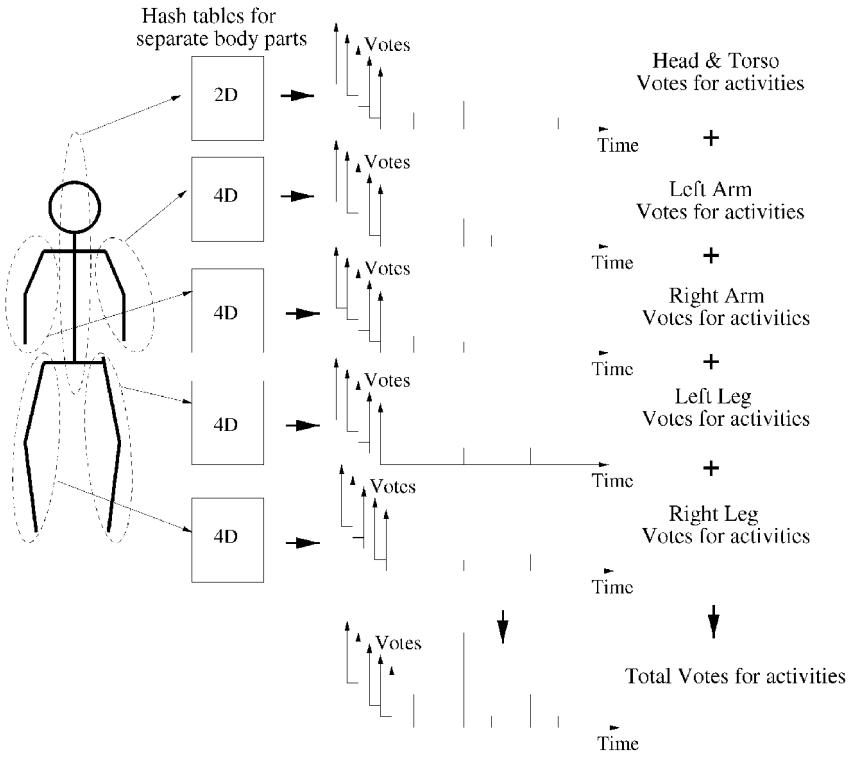
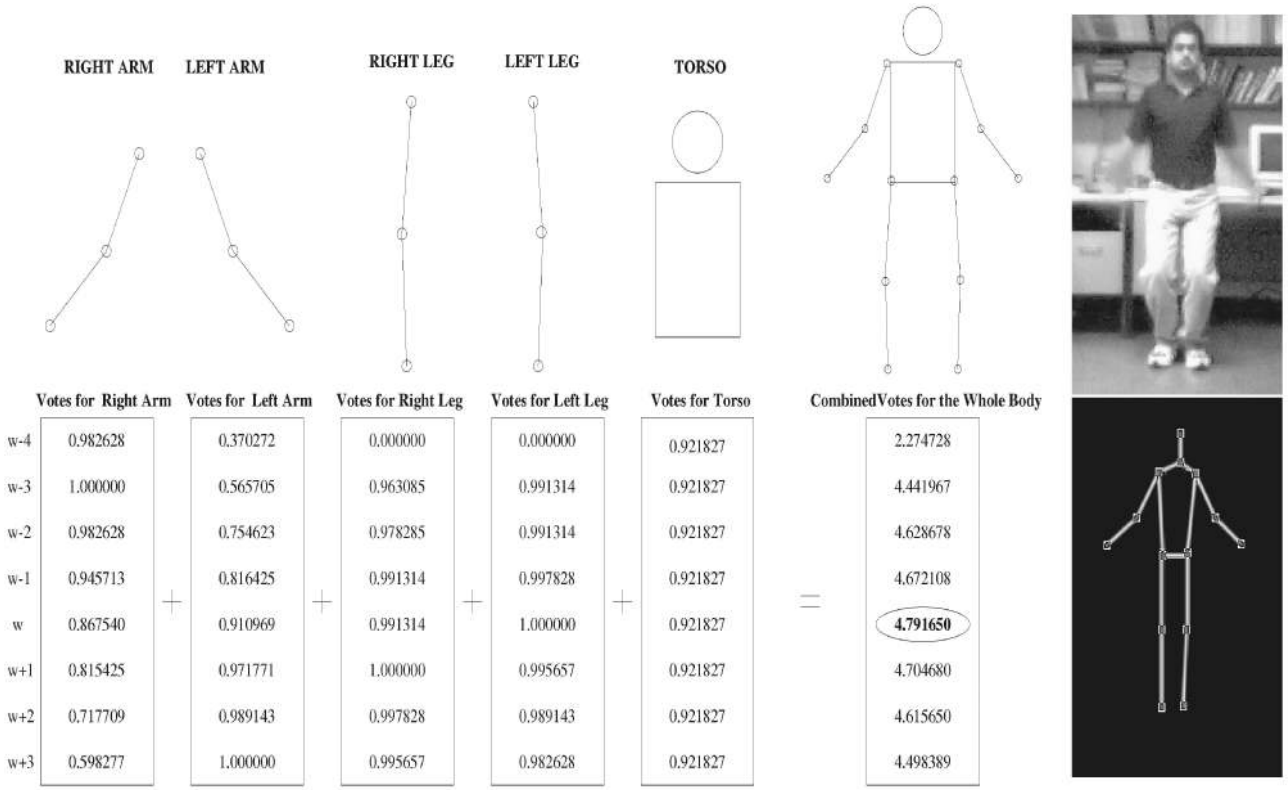


Fig. 9. A diagram of the whole voting process which illustrates how voting takes place for the different body parts for different model activities and the way in which the votes for different body parts are combined.



	Votes for Right Arm	Votes for Left Arm	Votes for Right Leg	Votes for Left Leg	Votes for Torso	Combined Votes for the Whole Body
w-4	0.982628	0.370272	0.000000	0.000000	0.921827	2.274728
w-3	1.000000	0.565705	0.963085	0.991314	0.921827	4.441967
w-2	0.982628	0.754623	0.978285	0.991314	0.921827	4.628678
w-1	0.945713	0.816425	0.991314	0.997828	0.921827	4.672108
w	0.867540	0.910969	0.991314	1.000000	0.921827	**4.791650**
w+1	0.815425	0.971771	1.000000	0.995657	0.921827	4.704680
w+2	0.717709	0.989143	0.997828	0.989143	0.921827	4.615650
w+3	0.598277	1.000000	0.995657	0.982628	0.921827	4.498389

Fig. 10. A numerical example illustrating the process of voting for the five body parts for a pose corresponding to a jumping test sequence. The votes shown are with respect to different frames of model jumping activity stored in the hash table. The votes are shown for a subset of model frames which include the winning pose and its neighbors. The frame numbers of the poses which are shown are represented in terms of the winning frame w. The test pose which is being voted for is shown in the picture in the top right-hand corner and the model pose which received the highest vote is shown in the bottom right-hand corner.

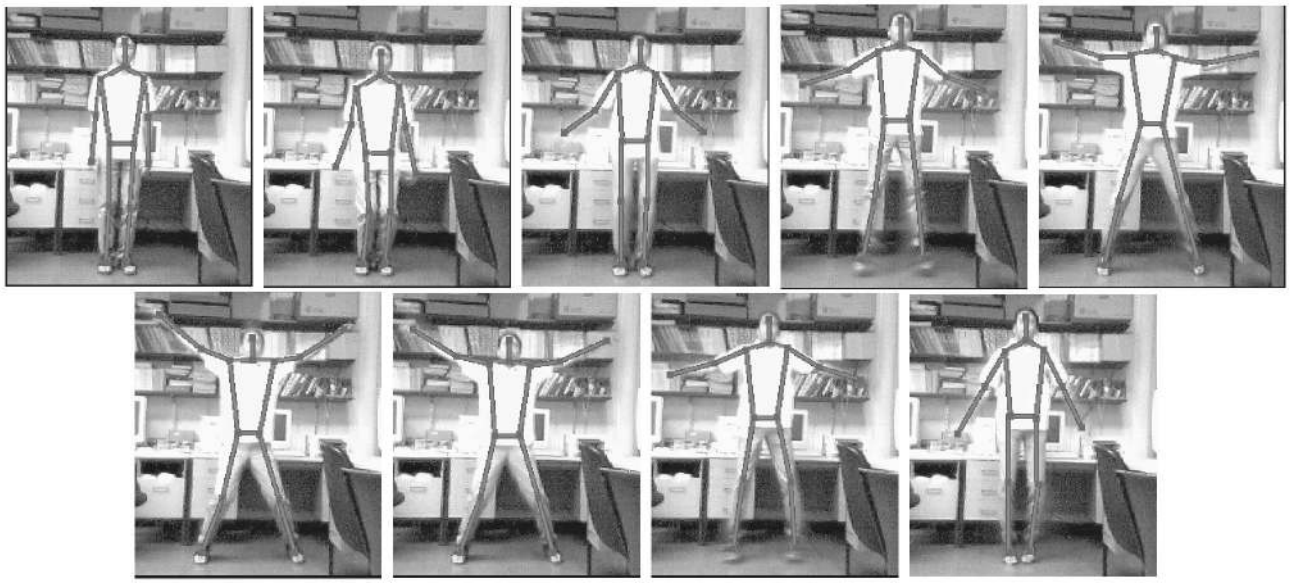


Fig. 11. Sample frames of jumping sequence to illustrate the full range of the activity. The skeleton superimposed on the human body represent the detected parts.

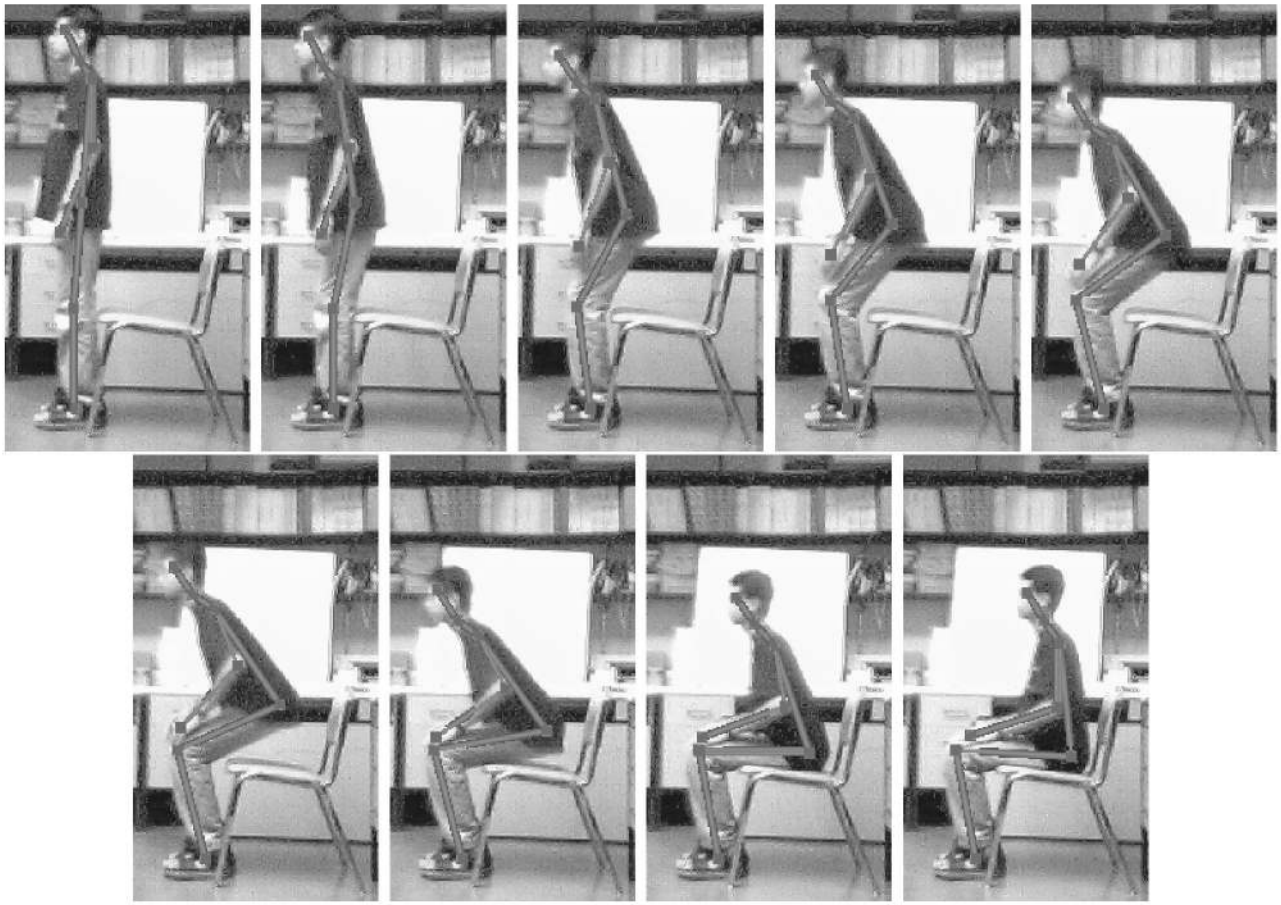


Fig. 12. Sample frames of sitting sequence to illustrate the full range of the activity. The skeleton superimposed on the human body represents the detected parts.

5 Experimental Results

In this work, we use a novel method for human body tracking using EXpansion Matching (EXM) [4]. However, since this paper is focused on a novel method for human activity recognition and since there are many other methods for human tracking (see a survey in [10]), we shall describe our method very briefly in the Appendix.

We applied our method to a database of eight different human activities. These activities are jumping, kneeling, picking up an object, putting down an object, running, sitting

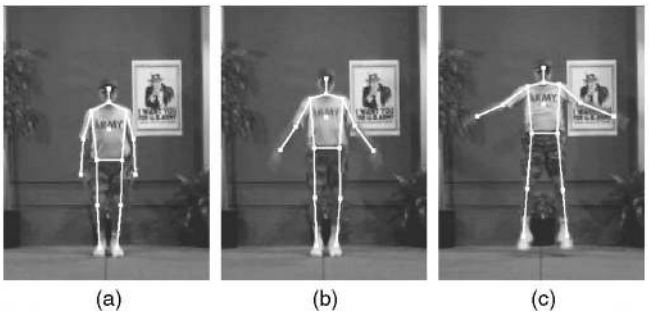


Fig. 13. The frames of a video stream of a man jumping overlaid with wire frames representing detected parts. (a) Frame 1, (b) frame 5, and (c) frame 9.

down, standing up, and walking. A total of 92 activity video sequences are captured in our lab using a Kodak digital video camera connected to a Compaq AMD K7 computer. Out of the 92 videos, we used 24 as model activities and 68 as test activities. Out of the 68 test activities, 28 videos are taken at seven views and are used for analysis of the view point sensitivity of our scheme and the remaining 40 videos are used as regular test activity, where the activities are performed by five people. The eight model activities are derived from 24 videos where each model is obtained after averaging over a sequence of three videos performed by three people. Figs. 11 and 12 show sample frames of the jumping and sitting activity. For each video recording, we first manually locate the position, direction, and size of each body part of the subject in the first frame of the video. Next, the EXM tracking method is used without manual intervention for the whole video stream. The tracking results are shown in Fig. 13 and Fig. 14, which are overlaid on the original images. Although we display, for each recording, only a few frames, the tracking is performed on successive frames in each video recording. Fig. 13 shows a man jumping up and down and Fig. 14 shows a person walking in a cluttered environment. From these results, we determine the 2D location of all the parts and store them in a database. Further, the 2D angles of all the body parts are determined and five different hash tables are created for the torso and the four limbs.

The experiment is conducted on a test set of 40 activity videos, which is different from the model set. The test set consists of five different people performing each of the eight

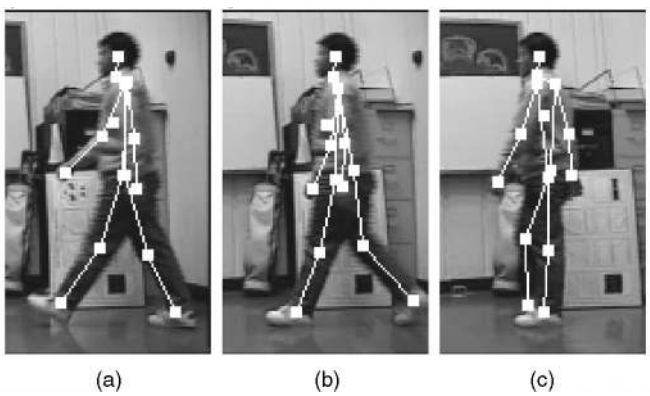


Fig. 14. The frames of a video stream of a man walking overlaid with wire frames representing detected parts. (a) Frame 5, (b) frame 8, and (c) frame 11.

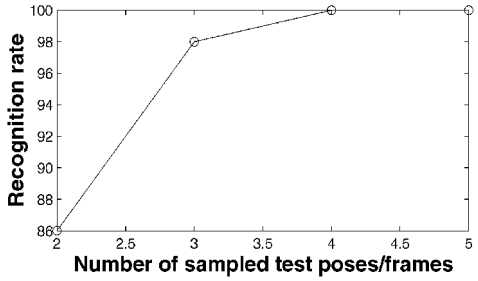


Fig. 15. A graph of the recognition rate versus number of sparse test frames used in the voting. One hundred percent recognition rates are achieved with only four sampled poses.

activities. The test frames are generated by taking four frames from each test set which are sampled at random time intervals. To ensure independence, we restrict the random samples to not being too close to one another.

To determine how many frame samples are required for robust recognition, we examine the "number of frames" versus the "recognition rate." As shown in Fig. 15 above four frames the recognition rate is close to 100 percent. Thus, we require a minimum of four frames for all our recognition experiments. These results conform with the probabilities for false matching derived in (7).

Table 1 displays the average votes for each possible activity pair for this method. It can be observed that the average score for each test activity is the highest for the correct model activity in the table. We also test each individual score and find that all of them are correct. This shows that the method recognizes all the activities correctly.

We also perform experiments to study the occlusion sensitivity of our recognition method in conditions of partial occlusion of some body parts. The results have been tabulated in Table 2. This demonstrates our method's inherent robustness to partial occlusion due to the independent structure of the five hash tables of the different body parts. The method maintains high correct recognition rates even when two body parts are occluded.

We also perform experiments to find the view point sensitivity of our recognition scheme with activities performed at azimuth angles ranging from 0-90 degrees. Fig. 16 shows an example of the different views of an activity tested. Fig. 17 shows the recognition results that we obtain for activities viewed at different angles. The results show that our method is invariant to view point variations to the extent of ± 30 degrees in azimuth.

6 Conclusions

Human activity recognition finds application in the fields of human-machine interaction, security surveillance, content-based retrieval, etc. In this paper, we present a novel method for view-based human activity recognition by indexing into a multidimensional hash table followed by a sequence-based voting scheme.

This method gives us 100 percent recognition with 40 test videos. To evaluate the effectiveness of our method quantitatively, we define the Average Discrimination Ratio (ADR) as the average of the ratios of the first maximum vote to the second maximum vote for each activity. The average ADR for all the activities is 2.15, which means that the

TABLE 1
Average Votes of Activity Sequences for the Voting/Sequencing with Angular Pose and Velocity-Based Voting

	Jump	Kneel	Pick	Put	Run	Sit	Stand	Walk
Jump	**12.31**	3.91	1.97	2.00	2.18	2.00	1.20	3.55
Kneel	4.90	**9.99**	3.20	2.77	2.20	2.18	2.40	3.80
Pick	0.67	2.00	**8.00**	2.40	1.97	1.90	3.80	1.36
Put	1.95	2.58	3.10	**8.37**	1.58	4.71	2.50	1.74
Run	2.00	2.25	1.40	1.70	**3.23**	1.40	1.50	2.90
Sit	1.36	1.73	3.00	4.20	0.90	**8.60**	3.40	1.60
Stand	0.00	0.55	2.34	1.86	1.23	3.50	**9.90**	0.63
Walk	3.40	3.18	1.97	1.60	2.61	1.50	1.16	**5.75**

The rows correspond to test activity, while the columns correspond to the model activities. The best score in each row is in boldface numerals. The method yields correct recognition since the scores along the diagonal are the highest in each row.

correct activity receives more than double the votes of the second highest vote.

Our representations of the angles are only 2D projections of the actual 3D angles. Hence, our representation is limited to a given view of the activity and, therefore, our scheme is view-based. However, we find that this representation is not very sensitive to changes in vantage point and the viewing direction can be changed in the range of ± 30 degrees without seriously affecting the recognition rate. Currently, we are working on an inverse kinematics-based method for the recovery of the 3D angles from 2D images. Initial successful results show that our activity recognition system can be extended to be view invariant with only additional preprocessing.

In summation, we propose here a representation for human action/activity which can accurately describe any complex human activity/action and develop a robust method for activity recognition and retrieval. The indexing approach also provides an efficient storage and retrieval of all the activities in a small set of hash tables. The number of activities can be increased just by adding sampled model activities to the hash tables. As our experiments demonstrate, the method is also robust to partial occlusion.

APPENDIX

HUMAN BODY PART TRACKING USING EXPANSION MATCHING TECHNIQUE

In this work, we use a novel approach for human body tracking using EXpansion Matching (EXM) [4], [25]. The EXM filter is an efficient template matching approach and provides good results since it relies on image features with medium/high frequency content such, as texture and edges, which are present everywhere in the image. An important advantage of EXM matching is that it is highly robust to partial occlusion [3], [4]. Each body part is represented in the form of an ellipse, as illustrated in Fig. 18. For tracking of the human body, we construct an EXM filter for each body part. Further, the tracking is performed by application of these EXM filters successively and searching for maximum response. Tracking robustness of our approach is improved by a momentum-based updating scheme of the tracking filters. There have been many works [23], [20], [6], [19], [26], [5], [15], [16], [14], [21], [13] in the area of human body tracking. Ju et al. [13] represented body

TABLE 2
Correct Recognition Rate under
Occluded Conditions of Body Parts

Non-Occluded Body Parts	Correct Recognition Rate
1 Arm and Torso	70%
1 Leg and Torso	70%
1 Arm and 1 Leg	87.5%
1 Arm, 1 Leg and Torso	92.5%
Arms and Torso	85%
Legs and Torso	80%
Arms, 1 Leg and Torso	97.5%
Legs, 1 Arm and Torso	95%

The method maintains high recognition rates even when two body parts are occluded.

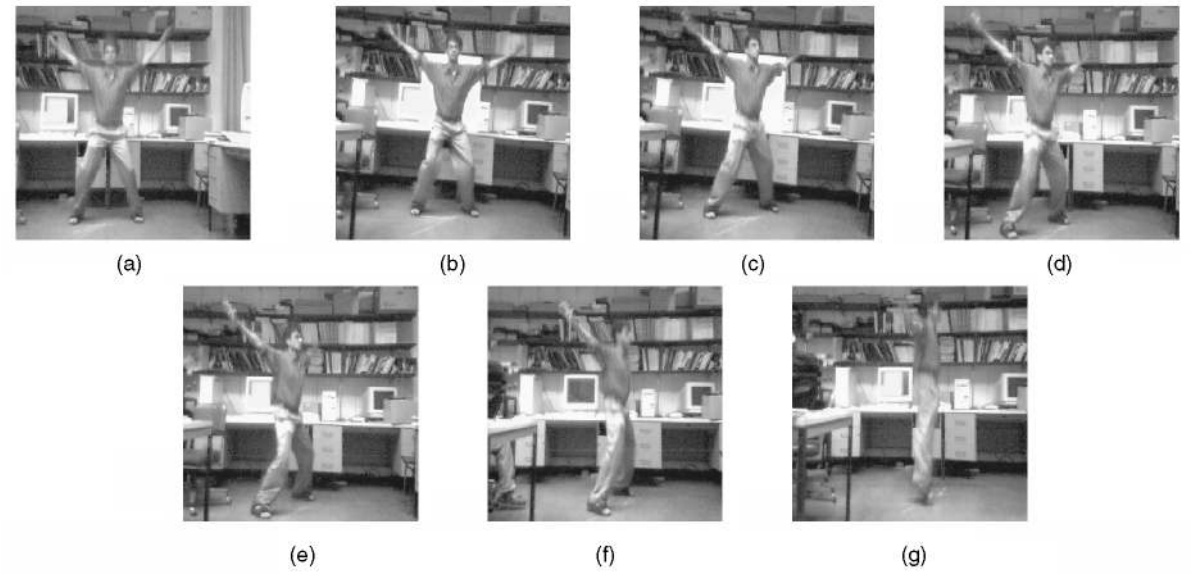


Fig. 16. Frames of jumping sequence for different views of the activity. The number on the lower left corner of each of the images represent the azimuth angle (in degrees) at which the activity is viewed. (a) Azimuth angle 0, (b) azimuth angle 15, (c) azimuth angle 30, (d) azimuth angle 45, (e) azimuth angle 60, (f) azimuth angle 75, and (g) azimuth angle 90.

parts in their tracking algorithm using patches and further used a parametric model for representing the motion. However, we choose to use our EXM-based method since it provides robust tracking even when the body part tracked is partially occluded or undergoes lighting variations.

The Expansion Matching approach is based on expanding the signal with respect to a set of basis functions that are all shifted versions of the template [4], [25]. In practice, the template, which serves as a basis function, is translated to all the candidate locations in the image. The magnitude of the expansion coefficients obtained at a particular location signifies the extent of the presence of the template at that location. Since EXM optimizes a novel Discriminative Signal to Noise Ratio (DSNR), which considers as unwanted clutter all the responses not at the center location, it achieves a very sharp output peak where the template matches the image. In contrast, the widely used matched filtering (correlation-based matching) usually outputs a very broad peak which is also more sensitive to occlusion.

For human body tracking, one needs to obtain a suitable set of templates for effective tracking. In this paper, we use a human model similar to the one used in [5], [16] with

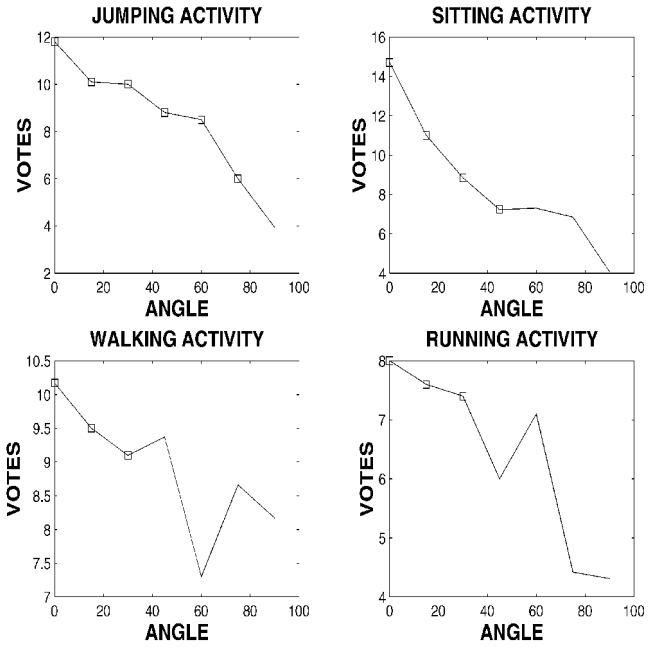


Fig. 17. A graph of the votes that were obtained for different view angles of four different activities (jumping, sitting, walking, and running activity). The squares in the graph represent correct recognition of activity. The results show that all the activities were correctly recognized at least up to 30 degrees.

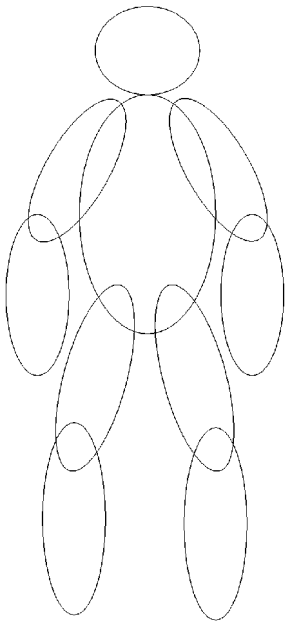


Fig. 18. A model of human body parts for EXM filtering.

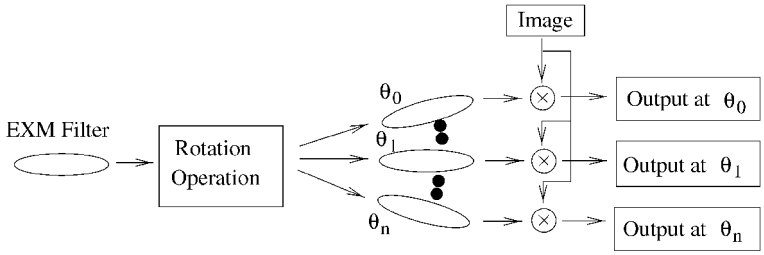


Fig. 19. The processing of 3D EXM filtering.

slight variations. The human body is represented by nine cylinders for the torso, upper arms and legs (thighs), and lower arms (forearm plus hand) and legs (calf plus foot), plus a sphere for the head. Each of the body parts is considered as a rigid object that can rotate and translate. In our 2D representation, the cylinders are represented as elliptical regions in the image, as described in Fig. 18. The representation of the posture of the body is composed of the poses of individual parts. In our convention, all the poses are relative to the positive x-axis. The pose of the torso is described by the position of its mass center and the direction of its major axis. The upper arms posture is represented by the angles between their axes and the positive x-axis and the posture of the lower arms is represented by their angles between their axes and the positive x-axis. Similar representation is used for the legs.

Here, we assume that the initial positions and sizes of these parts are known and our goal is to track these parts in a video stream. Since the extent of the possible relative motion between head and torso is very small compared with their sizes, we think that it is reasonable to designate only one EXM filter for these two body parts. For the rest of the body parts, we designate one filter for each. Hence, for tracking of the entire body, we have nine EXM filters.

The tracking is performed by applying the set of nine EXM filters represented in elliptical regions to each frame in the video stream. An example of EXM filtering is shown in Fig. 21 which shows detection of the elbow. To reduce computations, the filters are applied only in the vicinity of previous positions of these parts. The positions of the parts are updated sequentially. Since all parts move around the torso and, generally, upper arms and legs move less than lower arms and legs, the order of the application of the EXM filters corresponds to their movement relative to the torso. Obviously, we should use the torso-head filter first to find the new position of the major portion of the body in the next image. The motion of the torso possibly consists of translation, as in the case of walking, as well as rotation,as in the situation of bending. If the related translation and rotation are continuous over time, then one needs to apply the torso EXM filter in the neighborhood of the current position in the three-dimensional space (x, y, θ) with x and y representing translation and θ for rotation in the image plane. The best new position would be the one at which the output of the filter achieves the maximum in the (x, y, θ) space. These three-dimensional variations can be implemented by first rotating the EXM filter by a discrete set of directions and then applying the set of rotated filters to the image. The process of the three-dimensional filtering is depicted in Fig. 19.

Once the new position of the torso is found, the current positions for the upper arms and legs need to be updated before applying corresponding EXM filters. The positions of these upper arms and legs are represented by the coordinates of the four joints of the upper arms and legs with the torso, plus the angles formed between them and the positive x-axis. Under the assumption of smooth motion over time, one can use the same three-dimensional (x, y, θ) space as with the torso. This again requires rotating corresponding filters, applying them to the image, and searching for the point with maximum response. Upon the acquisition of new positions of four upper arms and legs, the current positions of the remaining four lower arms and legs need to be updated in a similar manner. The steps used to locate the new positions for all the parts of the body are illustrated in Fig. 20.

In some video sequences, one or more body parts may be occluded in a few video frames. Such an occlusion is readily detected by the low matching scores. The body part usually reappears after a few frames and our tracking filters usually find it. For such cases, we use an interpolation scheme that determines the angular pose of such body parts in the frames in which they are occluded.

After a complete round of filtering and searching, the set of part EXM filters need to be updated to accommodate the new positions, orientations and lighting variations. All these factors may change the appearance of the body part and, hence, the template. We employ here a momentum approach for these updates. The momentum formula keeps a portion α of the past template and combines it with the new template. According to our experiments, the best results are achieved when $\alpha = 0.9$. When α is too small, there is a risk that a temporary imaging noise may cause a

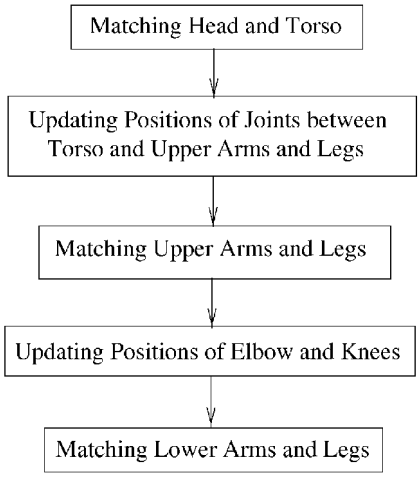


Fig. 20. The order of 3D EXM filtering of body parts (top-down).

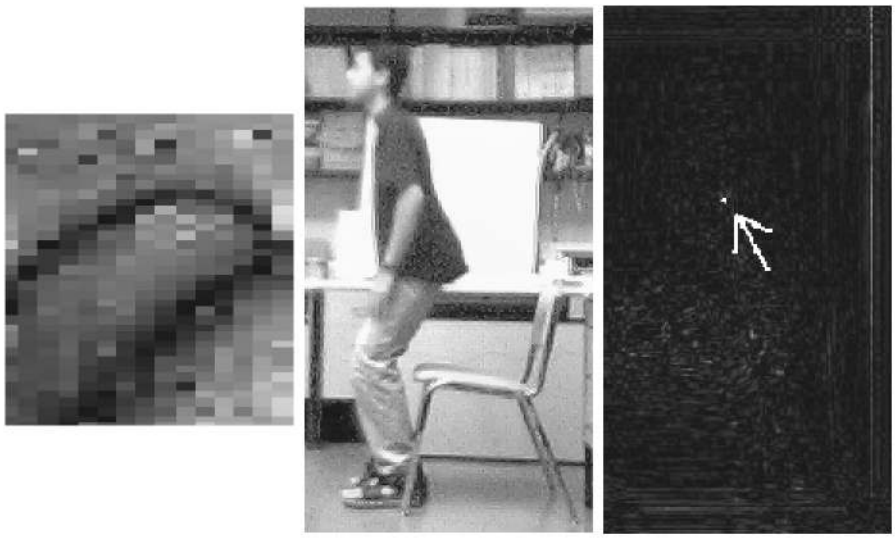


Fig. 21. An example of the detection of an elbow using EXM. (a) Shows the filter which corresponds to the elbow, (b) is a frame of the sitting sequence, and (c) is the result of application of the EXM filter (a) on the image in (b). Please note the strong peak found (marked by the arrow).

complete loss of tracking. The updated template $f_n(x,y)$ is formed from the old template $f_o(x,y)$ and parts in the EXM processed image $i(x,y)$ as

$$f_n(x,y) = \alpha \cdot f_o(x,y) + (1-\alpha) \cdot i(x,y), \qquad (15)$$

where α is a constant and determines the momentum.

ACKNOWLEDGMENTS

This work was supported by US National Science Foundation (NSF) grants nos. IIS-9711925, IIS-9876904 and IIS-9979774.

REFERENCES

[1] A.F. Bobick and J.W. Davis, "An Appearance Based Representation of Action," *Proc. 13th Int'l Conf. Pattern Recognition,* Aug. 1996.
[2] A.F. Bobick and J.W. Davis, "The Recognition of Human Movement Using Temporal Templates," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 3, Mar. 2001.
[3] J. Ben-Arie and K.R. Rao, "A Novel Approach for Template Matching by Non-Orthogonal Image Expansion," *IEEE Trans. Circuits and Systems for Video Technology,* vol. 3, no. 1, pp. 71-84, Feb. 1993.
[4] J. Ben-Arie and K. R. Rao, "Optimal Template Matching by Non-Orthogonal Image Expansion Using Restoration," *Int'l J. Machine Vision and Applications,* vol. 7, no. 2, pp. 69-81, Mar. 1994.
[5] E. Di Bernardo, L. Goncalves, and P. Perona, "Monocular Tracking of the Human Arm in 3D: Real-Time Implementation and Experiments," *Proc. Int'l Conf. Pattern Recognition,* pp. 622-626, Aug. 1996.
[6] M. La Cascia, J. Isidoro, and S. Sclaroff, "Head Tracking via Robust Registration in Texture Map Images," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR-98),* pp. 508-514, June 1998.
[7] C. Barron and I. A. Kakadiaris, "Estimation Anthropometry and Pose from a Single Image," *Proc. Conf. Computer Vision and Pattern Recognition,* pp. 669-676, June 2000.
[8] H. Fujiyoshi and A.J. Lipton, "Real-Time Human Motion Analysis by Image Skeletonization," *Proc. Workshop Application of Computer Vision,* Oct. 1998.
[9] A. Galata, N. Johnson, and D. Hogg, "Learning Variable-Length Markov Models of Behaviour," *Computer Vision and Image Understanding,* vol. 81, no. 3, pp. 398-413, Mar. 2001.
[10] D.M. Gavrila, "The Visual Analysis of Human Movement: A Survey," *Computer Vision and Image Understanding,* vol. 73, no. 1, pp. 82-98, 1999.
[11] I. Haritaoglu, D. Harwood, and L.S. Davis, "w^4: Real-Time Surveillance of People and Their Activities" *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 8, pp. 809-830, Aug. 2000.
[12] Y.A. Ivanov and A.F. Bobick, "Recognition of Visual Activities and Interactions by Stochastic Parsing," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 8, pp. 852-871, Aug. 2000.
[13] S.X. Ju, M.J. Black, and Y. Yacoob, "Cardboard People: A Parameterized Model of Articulated Image Motion," *Proc. Second Int'l Conf. Automatic Face- and Gesture-Recognition,* pp. 38-44, Oct. 1996.
[14] M.K. Leung and Y.H. Yang, "First Sight: A Human Body Outline Labeling System," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 17, no. 4, Apr. 1995.
[15] A. Lipton, H. Fujiyoshi, and R. Patil, "Moving Target Detection and Classification from Real-Time Video," *Proc. 1998 DARPA Image Understanding Workshop (IUW '98),* Nov. 1998.
[16] D. Marr and H.K. Nishihara, "Representation and Recognition of the Spatial Organization of Three-Dimensional Shapes," *Proc. Royal Soc. London B,* vol. 200, pp. 269-294, 1978.
[17] T.B. Moeslund and E. Granum, "A Survey of Computer Vision-Based Human Motion Capture," *Computer Vision and Image Understanding,* vol. 81, no. 3, pp. 231-268, Mar. 2001.
[18] T.K. Moon and W.C. Stirling, *Mathematical Methods and Algorithms for Signal Processing.* Prentice Hall, 2000.
[19] A. Pentland and B. Horowitz, "Recovery of Non-Rigid Motion and Structure," *IEEE Trans. Pattern Anaysis and Machine Intelligence,* vol. 13, no. 7, pp. 730-742, July 1991.
[20] J.M. Rehg and T. Kanade, "Model-Based Tracking of Self-Occluding Articulated Objects," *Proc. Fifth Int'l Conf. Computer Vision,* pp. 612-617, 1995.
[21] K. Rohr, "Towards Model Based Recognition of Human Movements in Image Sequences," *CVGIP: Image Understanding,* vol. 59, 1994.
[22] R. Polana and R. Nelson, "Recognizing Activities," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 815-818, 1994.
[23] N. Sawasaki, T. Morita, and T. Uchiyama, "Design and Implementation of High-Speed Visual Tracking System for Real-Time Motion Analysis," *Proc. Int'l Conf. Pattern Recognition,* pp. 478-484, 1996.
[24] J. Schlenzig, E. Hunter, and R. Jain, "Vision Based Hand Gesture Interpretation Using Recursive Estimation," *Proc. 28th Asilmoar Conf. Signals, Systems, and Computers,* 1994.
[25] Z. Wang and J. Ben-Arie, "Optimal Ramp Edge Detection Using Expansion Matching," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 11, pp. 1092-1098, Nov. 1996.

[26] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-Time Tracking of Human Body," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 780-785, July 1997.
[27] J. Yamato, J. Ohya, and K. Ishii, "Recognizing Human Action in Time-Sequential Images Using Hidden Markov Model," *Proc. Conf. Computer Vision and Pattern Recognition,* pp. 379-385, June 1992.
[28] M.-H. Yang and N. Ahuja, "Recognizing Hand Gesture Using Motion Trajectories," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 466-472, June 1999.

Jezekiel Ben-Arie received the BSc, MSc, and DrSc degrees from the Technion, Israel Institute of Technology, Haifa. Currently, he is a professor with the Electrical and Computer Engineering Department, University of Illinois, Chicago. His areas of specialization are machine vision, image processing, and neural networks. He has also worked in auditory processing and sound localization. In these areas, he has more than 120 technical publications. In 1986, he discovered the probabilistic peaking effect of viewed angles and distances. In 1992, he developed the nonorthogonal expansion matching (EXM) method, which is very useful in the recognition of highly occluded objects and in motion video compression. More recently, he developed the Affine Invariant Spectral Signatures (AISS) and the Volumetric Frequency Representation (VFR) for 3D shape representation and recognition and a novel approach for human activity recognition by indexing-sequencing. Dr. Ben-Arie currently serves as an associate editor of the *IEEE Transactions on Image Processing*. He is a member of the IEEE and the IEEE Computer Society.

Zhiqian Wang received the BE degree in electrical engineering from the University of Science and Technology of China, Hefei, in 1984, the MS degree in electrical engineering from Tsinghua University, Beijing, China, in 1987, and the PhD degree in electrical engineering and computer science from the University of Illinois, Chicago, in 2000. Currently, he is with RCT Systems Inc., Chicago. His research interests are in signal and image processing, neural networks, and image understanding. He has authored or coauthored more than 30 publications in these areas. Dr. Wang is a member of Sigma Xi. He was the recipient of the Outstanding PhD Thesis Award of 2000 at the University of Illinois at Chicago. He is a member of the IEEE.

Purvin Pandit received the BS degree in eletronics engneering (with distinction) from the University of Mumbai, India, in 1999 and the MS degree in electrical engineering from the University of Illinois at Chicago in 2001. He is a member of the technical staff with Thomson Multimedia Corporate Research where he is working on compression technologies for mobile applications in 3D CDMA technology. He has published papers on the topics of computer vision. His professional interests include transmission of digital compressed video over wireless links, digital video compression technologies with a focus on optimization, error concealment, and transform coding. He is an active member of the IEEE-Signal Processing chapter. He is a member of the IEEE.

Shyamsundar Rajaram received the BS degree in electrical and electronics engineering from the University of Madras in 2000. He is currently completing the MS degree in electrical and computer engineering at the University of Illinois at Chicago. He has many papers in the field of computer vision. His research interests are in the fields of computer vision, signal and image processing, and computational geometry. He is a student member of the IEEE.