

REAL TIME HUMAN ACTIVITY IDENTIFICATION SYSTEM TO AID ELDERLY CARE

Udaykiran.K¹, H.Chidananda², Ramya BR³

¹B.E. 8th semester, Department of computer science and engineering, RYMEC, Karnataka, India.

²Department of computer science and engineering, RYMEC, Karnataka, India.

³B.E. 8th semester, Department of computer science and engineering, RYMEC, Karnataka, India

Abstract-Human activity recognition has been a tremendous growth in the last decade playing a major role in the field of pervasive computing. This emerging popularity can be attributed to its myriad of real life applications primarily dealing with human-centric problems like healthcare and eldercare. Many research attempts with data mining and machine learning techniques have been undergoing to accurately recognize human activities for e-health systems. Aim of this project is to recognize the activity under care using a cost effective camera rather than multiple costly sensors by using adaptive background subtraction for removal of the background, optical flow model for feature extraction and iterative querying heuristic (IQH) and multiple instance learning model for recognizing the activities and to send a sms using GSM module to the care takers in real time.

Keywords: Human Activity, Iterative Querying Heuristic (IQH) optimization algorithm, healthcare, eldercare, Multiple Instance Learning Algorithm.

1. INTRODUCTION

Human activity identification based on pc vision is of great scientific and practical importance. Its various important theoretic values and wide potential applications, like real-time video surveillance system, interpretation of sport events, and human pc interactions. In recent years, human activity identification has become a search hot spot. Tremendous amount of researches had been made in the field of human activity identification from video sequence. There are many human activity identification strategies are already present. Common methodology to identify human activity aims at both still images and video sequences.

The human activity recognition continues to be an immature technology and a challenging problem in pc vision to attain robust human activity identification from image sequences due to occlusion, human postures, illumination conditions, dynamic backgrounds, camera movements. With moving target and cameras, non-stationary background, few vision algorithms will categorize and recognize human activity well. It's essential

for intelligent and natural human pc interaction to recognize human actions automatically.

In this paper, we tend to present a unique technique for human activity recognition from video sequences. This approach includes two steps of feature extraction and human activity identification .In the extracting feature, global silhouette features of human are extracted by using adaptative background subtraction algorithm; local optical flow features are extracted by using optical flow model. Then fusing global silhouette feature vector and local optical flow feature vector form a hybrid feature vector.

Finally an optimized Multiple Instance Learning algorithmic program is used for human activity recognition. Additionally, in order to improve the recognition accuracy, an Iterative Querying Heuristic (IQH) optimization algorithmic program is employed to train the Multiple Instance Learning model.

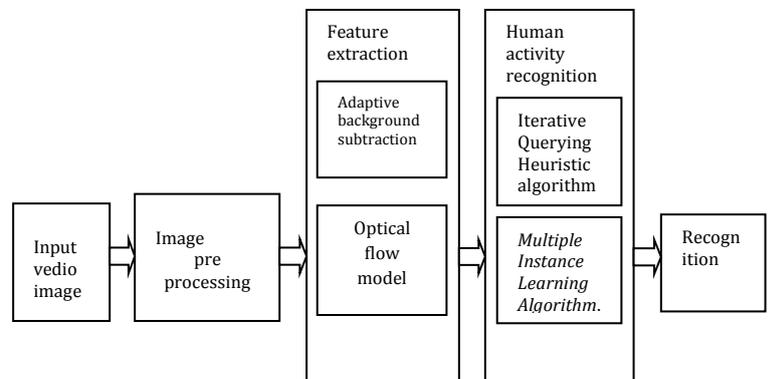


Figure1: Flow diagram of human activity reorganization based on fusion features extraction.

Given unlabeled human activity video sequence, it's our goal to automatically learn totally different human activity categories in the data and apply the Multiple Instance Learning model for human actions categorization and recognition in the new video sequences. Our approach is illustrated in Figure 1.

The rest of this paper is organized as follows. In Section 3, we tend to describe human activity features presentation based on adaptive background subtraction and optical flow models. In Section 4, we tend to explain about the algorithmic rule for human activity identification using an

optimized Multiple Instance Learning algorithmic program. Section 5 shows experimental results, and also the conclusions are given in the final section.

2. Existing system

Vision based strategies are often used with none restrictions upon the user. Hand tracking is a significant section for gesture recognition, within which the signer's or user's hands should be detected and localized in image frames. Wherever detection needs correct lightening if not the gesture can't be captured accurately.

3. Proposed System

The method of identifying a specific activity with the help of camera, the proper preprocessing of their provided info and therefore the learning/reasoning using this info. If choosing of the sensors and the data processing strategies are wrongly performed, the entire activity detection method could fail, resulting in the consequent failure of the entire application. The target of this project is to classify the most activities considered in smart home situation that are with respective to elderly people's independent living, as well as their characterization and formalized context presentation and is to assist researchers and developers in these lower-level technical aspects that are nevertheless fundamental for the success of the entire application.

3.1 Human Activity Features Illustration

Recognizing human activities from video sequences is both a difficult task and an interesting experimenting area. Generally, two vital queries are concerned in human activity recognition. One way is to represent expeditiously human activity, that is the key for a human activity recognition methodology, and also the other way is to model reference movements, which expeditiously cope with variations at spatial and temporal scales among similar motion categories.

Extracting helpful motion info from raw video information is crucial for human activity recognition. The selection of the motion features affects the results of the human activity recognition methodology directly. Several factors usually influence the only feature differently, such as appearance of human body, atmosphere, and video camera. Therefore the accuracy of action recognition is restricted.

After completely considering the benefits and downsides of various features, we tend to propose a hybrid feature methodology on the idea of studying the representation and recognition of human activity that is fusing global

silhouette feature and local optical flow feature.

3.1.1. Global Silhouette Feature illustration.

Human silhouettes in single frame image can be used to describe the data of which overall shape of a human body movement changes. Adaptive background subtraction could be a quite excellent methodology for global silhouette feature extraction. We tend to use it to work out the general area of the movement and the human body shadow. Assume that every actions that are performed before the static background. We tend to use the adaptive background subtraction to determine the motion area and extract the human body silhouette.

The step for adaptative background subtraction algorithmic rule is as follows.

Step 1 (initialize the background model). Initially Taking T background image continuously, then through the image, a single Gaussian distribution is created to set the initial background statistical model $H(u_i, v_i^2)$. Forming the Gaussian distribution needs the mean and standard deviation, which might be calculated using

$$u_i = \frac{1}{T} \sum_{t=1}^T u_{it} \quad (1)$$

$$v_i^2 = \frac{1}{T} \sum_{t=1}^T (u_{it} - u_i)^2$$

Where u_{ij} is color value of the point i in the t^{th} image.

Step 2 (extracting foreground area). Assumptive that color value of the point i is in current image, the image binarization is expressed as

$$B_i = \begin{cases} 1, & \text{if } (r_i - u_i) > 3\sigma_i \\ 0, & \text{else,} \end{cases} \quad (2)$$

Where points of all signs "1" are foreground area and points of all signs "0" are background region.

Step 3 (update background model). Assume that (t) is color expectations of the point i at time t , $V_i^2(t)$ is color variance of the point i at time t , and $r_i(t)$ is color value of the point i in time t which collect pictures. At time $t + 1$, there is

$$u_i(t + 1) = \begin{cases} \alpha u_i(t) + (1 - \alpha)r_i(t), & (\text{flag}_i = 0) \\ u_i(t), & (\text{flag}_i = 1), \end{cases}$$

$$v_i^2(t + 1) = \begin{cases} \alpha v_i^2(t) + (1 - \alpha)(r_i(t) - u_i(t))^2 & (\text{flag}_i = 0) \\ v_i^2(t), & (\text{flag}_i = 1) \end{cases} \quad (3)$$

Where α is used to control the background update rate, $0 < \alpha < 1$. Each point in the binary image is marked by flag i , if value of the flag i is "1,"that is foreground, and flag i is "0,"that is the background. Thus, the area marked as "0" constitutes the background region and the area marked as "1" constitutes the foreground area.

So we will get the foreground of the binary image. Figure 2 shows an example of adaptive background subtraction to extract the moving human body silhouette. Silhouette features have the subsequent benefits. Silhouette features can be used to describe the shape of human movement information simply and visually. Silhouette features are easy to be extracted.

Binary silhouette figure isn't sensitive to texture and color of the foreground image. Suppose that a video sequence contains T frame image I —namely, $V = [I_1, 2, \dots, I_T]$ —and S is the sequence of motion silhouette corresponding to video V —namely, $S = [s_1, s_2, \dots, s_T]$. We tend to use the contour vector to explain the overall human silhouette and shape info.

The step is as follows.

Step 1. Use canny operator for edge profile of every frame silhouette and calculate the coordinates of the edge profile. Such human body contour can be set with n points, namely, $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$.

Step 2. Calculate the center of mass regarding human body contour, which is expressed as

$$(x_c, y_c) = \left(\frac{1}{n_t} \sum_{i=1}^{n_t} x_i, \frac{1}{n_t} \sum_{i=1}^{n_t} y_i \right) \quad (4)$$

Where (x_c, y_c) is the center of mass and (x_i, y_i) is edge points of the contour. n_t is the number of edge points in the t^{th} image.

Step 3. Calculate the distance from the center to the edge points, that is expressed as

$$d_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}, \quad i = 1, 2, 3, \dots, n_t \quad (5)$$

Where d_i is the edge distance from the center of mass to edge points, which the i^{th} edge points correspond to.

Step 4. (Normalize contour vector). We tend to use 2-norm for contour vector normalization process. 2-norm is the most popular general approaches for computing approximations of the condition number. It's the incremental condition estimation (ICE) of a triangular matrix and might be viewed as a special case of the framework. The strategy is of course connected to adaptive techniques and contains clearly visible links to matrix decompositions. After the normalization process of contour vector, we perform interval such as resampling. We obtain the fixed point N , where N is set as 180.

3.1.2. Local Optical Flow Feature illustration.

Silhouette image extraction of incorrect information might lead to the contour vector characteristics which cannot accurately express movement characteristic. Optical flow characteristics are effective and accurate motion info representation in video sequences. Optical flow-based human actions representation has attracted much attention. The human activity could be a dynamic event; it can be represented by the motion info of the human. The

local optical flow features (optical flow vector) are estimated by optical flow model.

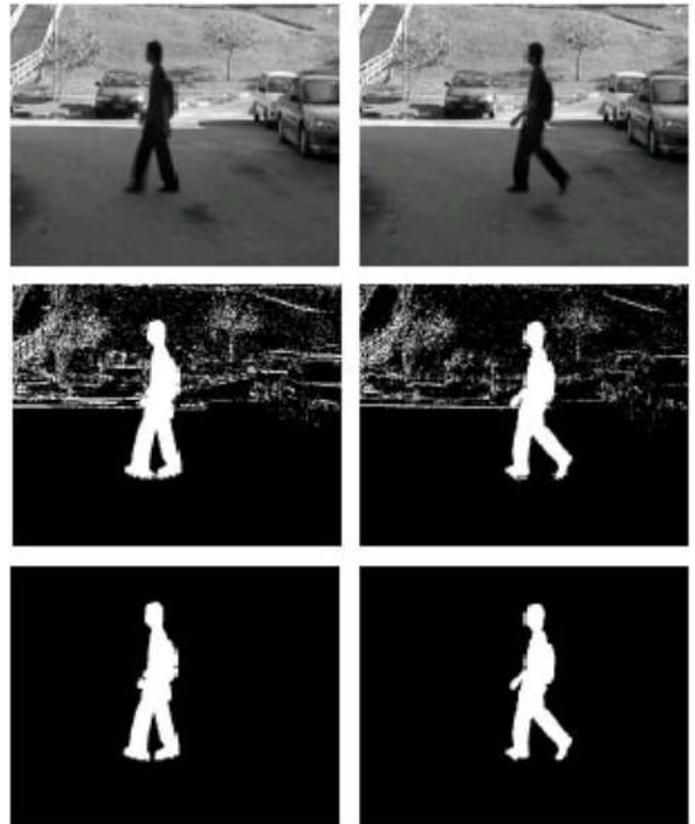


Figure 2: The processing pipeline of the adaptive background subtraction: (top) the original picture, (middle) the object extraction background subtraction and, and (bottom) the body silhouettes

Firstly, we tend to calculate the optical flow vector of human actions, namely $\mu = (\mu_L, \mu_H)$ at every frame using optical flow equation, that is expressed as

$$I_L \mu_L + I_H \mu_H + I_t = 0, \quad (6)$$

Where $I_L = \frac{\partial I}{\partial x}, I_H = \frac{\partial I}{\partial y}, I_t = \frac{\partial I}{\partial t}, \mu_L = \frac{dx}{dt}$ and $\mu_H = \frac{dy}{dt}$, where (x, y, t) is the image in the pixel (x, y) at time I , where $I(x, y, t)$ is the intensity in picture element (x, y) at time t , and μ_L, μ_H are the horizontal and vertical velocities in pixel (x, y) .

We can acquire $\mu = (\mu_L, \mu_H)$ by minimizing the objective function:

$$L = \int_D [\lambda^2 \|\nabla \mu\|^2 + (\nabla I \cdot \mu + I_t)^2] dx dy \quad (7)$$

We use the iterative algorithm to calculate the optical flow; optical flow is decomposed into longitudinal and transverse two components, namely, the longitudinal optical flow and transversal optical flow, that is expressed as

$$\begin{aligned} \mu_L^{k+1} &= \mu_L^{-k} - \frac{I_L [I_L \mu_L^{-k} + I_H \mu_H^{-k} + I_t]}{\lambda + I_L^2 + I_H^2} \\ \mu_H^{k+1} &= \mu_H^{-k} - \frac{I_H [I_L \mu_L^{-k} + I_H \mu_H^{-k} + I_t]}{\lambda + I_L^2 + I_H^2} \end{aligned} \quad (8)$$

Where k is the number of iterations, μ_L^{-k}, μ_H^{-k} are the average velocity of the neighborhood of point (x,y) , and $\mu_L^0 = \mu_H^0 = 0$ is initial value of velocity.

After standardization, optical flow diagrams are divided into 2×2 sub frame, where they are set $S1, S2, S3, S4$, respectively. The dimension of sub frame is set 60×60 , and the centers of sub frame are set $Gias$ shown in Figure 3, where $i = 1, 2, 3, 4$. Then each sub frame is split into 18 subareas in the centers of sub frame and degree of each central angle is 20° as shown in Figure 4, where they're set S_j^i and where $j = 1, 2, \dots, 18$.

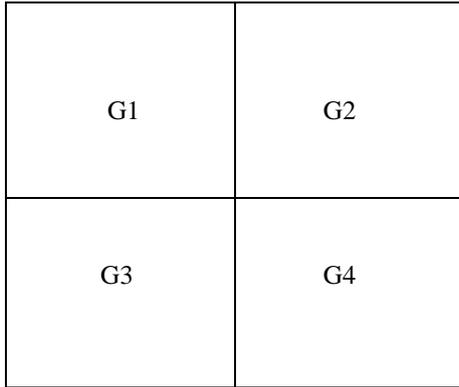


Figure 3: 2×2 sub frame figure of optical flow diagrams.

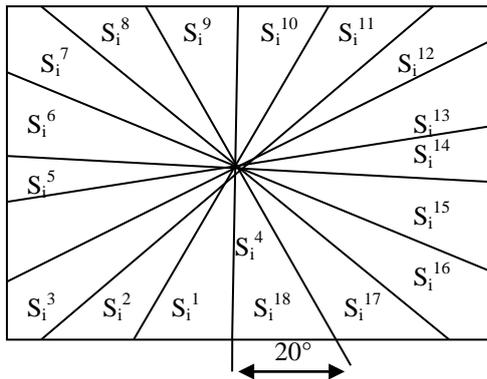


Figure 4: 18 subareas of each sub frame in the centers of sub frame.

Thus each frame image of optical flow graph is split into 72 subareas.

We calculate the sum of the longitudinal optical flow O_{Lij} , and the sum of the transverse optical flow O_{Hij} , at subarea:

$$O_{Lij} = \sum_{k=1}^K \mu_L^{k+1} \in S_i^j, \tag{9}$$

$$O_{Hij} = \sum_{k=1}^K \mu_H^{k+1} \in S_i^j,$$

where S_i^j is the subarea, $i = 1, 2, 3, 4$ and $j = 1, 2, \dots, 18$, and k is the number of the longitudinal optical flows or the transverse optical flows. Optical flow information of the image can be represented by the sum of longitudinal

optical flow and transverse optical flow in 72 subareas, which is expressed as

$$O_L = [O_{L_{1,1}, \dots}, O_{L_{1,18}, \dots}, O_{L_{4,1}, \dots}, O_{L_{4,18}, \dots}], \tag{10}$$

$$O_H = [O_{H_{1,1}, \dots}, O_{H_{1,18}, \dots}, O_{H_{4,1}, \dots}, O_{H_{4,18}, \dots}],$$

Where O_L is the sum of longitudinal optical flow in 72 subareas and O_H is the sum of transversal optical flow in 72 subareas. We calculate local optical flow vector O_T :

$$O_T = [O_L, O_H], \tag{11}$$

Where we use 2-norm for O_T normalization process and get the local optical flow vector of the current frame image.

3.1.3. Human Activity Hybrid Feature presentation.

In order to enhance the accuracy of human activity recognition, fusing the contour vector and local optical flow vector forma hybrid feature vector, which is expressed as

$$F_T = [O_T, D_T], \tag{12}$$

Where F_T is the hybrid feature vector of each frame image, O_T is local optical flow vector, and D_T is contour vector. Human actions are presented by the hybrid feature vectors of contour vector and local optical flow vector. As a result human actions may be regarded as motion, the local optical flow features will describe human actions effectively. Additionally, silhouette features will describe the shape of human movement information simply and visually. Thus, the hybrid features are shown to perform reliably with noisy image sequences and are applied in various tasks, such as action classification and action recognition.

4. Human Activity Identification

After characterizing human activities, there are several strategies to identify human activities. Because human actions recognition is innately a Multiple Instance Learning problem, we use the Multiple Instance Learning algorithmic rule to learn and recognize human actions. The Multiple Instance Learning model has been applied to numerous pc visions, such as object recognition, action recognition, and human detection. so as to extend recognition potency while not compromising with the accuracy, an Iterative Querying Heuristic (IQH) optimization algorithmic rule is employed to train the Multiple Instance Learning model, and later the improved Multiple Instance Learning framework is used to learn a unified classifier rather than of individual classifiers for all categories.

4.1. The Multiple Instance Learning Rule Analysis.

The Multiple Instance Learning is one among the foremost efficient machine learning algorithms nowadays. This idea

is known as an Integrated Segmentation and Recognition (ISR), and its key idea is to produce a special means in constituting training samples. Training samples are in “bags,” they’re not singletons, and all of the samples in a bag share a label. Every bag contains a large variety of instances. Samples are organized into positive bags or negative bags of instances, where a minimum of one instance is positive in a positive bag, however all instances are negative in a negative bag. In the Multiple Instance Learning, it should learn at the same time the samples in the positive bags are positive along with the parameters of the classifier. In this paper, the Multiple Instance Learning is employed for human actions with nonaligned training samples.

The step for recognizing human actions based on the Multiple Instance Learning is as follows.

Step 1. Given dataset $\{X_i, C_i\}_{N_i=1}$, where $X_i = \{x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{iN}\}$ is on behalf of training bags, that contains at least one positive sample in a positive bag, $C_i = \max_j(C_{ij})$ represents the score of the sample, and $C_i \in \{0, 1\}$, where $C_i = \{C_{i1}, C_{i2}, \dots, C_{ij}, \dots, C_{iN}\}$. N is the number of all weak classifiers.

Step 2. Update all weak classifiers k with the info $\{x_{ij}, C_i\}$.

Step 3. Initialize all strong classifiers: $= 0$ for all i, j , where strong classifiers are composed of all the weak classifier.

Step 4. For $k = 1$ to K , for $m = 1$ to N .

In the i^{th} bag, when the j^{th} sample is positive, the probability is calculated as follows:

$$P_{ij}^m = \sigma(H_{ij} + h_m(x_{ij})), \tag{13}$$

Where $P_{ij}^m = p(C_i | x_{ij}) = 1 / (1 + \exp(-C_{ij}))$,

In the positive bag, the probability is calculated as follows:

$$P_i^m = 1 - \prod_j (1 - P_{ij}^m), \tag{14}$$

Where $P_i^m = P(C_i / X_i)$,

The probability that is allotted to a group of training bags is expressed as

$$L^m = \sum_i (C_i \log(p_i^m) + (1 - C_i) \log(1 - P_i^m)) \tag{15}$$

End for.

Find the atmost of m^* from N and obtain the current optimal weak classifier as follows:

$$m^* = \arg_m \max(L^m), \tag{16}$$

The strong classifier is produced by the m^* as follows:

$$\begin{aligned} h_k(x) &\leftarrow h_{m^*}(x), \\ H_{ij} &= H_{ij} + h_k(x). \end{aligned} \tag{17}$$

End for.

Step 5. K weak classifiers constitute a strong classifier, which is expressed as

$$H(x) = \sum_k h_k(x). \tag{18}$$

Where h_k could be a weak classifier, which can create binary predictions by using $\text{sign}(H(x))$. In the Multiple Instance Learning, samples come into positive bags and negative bags of instances. Every instance x_{ij} is indexed by two indices, where i stands for the bag and j stands for the instance in the bag. All instances in a bag share a bag label C_i . The weight of every sample consists of the weight of bags and the weight of samples in the bag, where the amount of the samples can be interpreted as a likelihood ratio. P_{ij}^m is the probability of positive instances in the bags, so the weight of samples is P^m . We calculate $w_{ij} = \partial \log L^m / \partial y_{ij}$ and obtain the weight of the bags w_{ij} .

4.2. Optimizing the Multiple Instance Learning Algorithmic rules for Human Activity identification.

The Multiple Instance Learning algorithmic rules present a general paradigm for a more relaxed form of supervised learning. In the Multiple Instance Learning, the learner gets unordered sets of occurrences or baggage rather than receiving example or label pairs, and labels are provided for every baggage instead for each occurrence, where a positive baggage contains a minimum of one positive occurrence. In the initial training stages, training and evaluating have a direct impact on both the features and the appropriate thresholds selected, and it’s the key to a fast and effective classifier. The samples have high score in positive baggage. The final classifier labels these samples to be positive. The remaining samples have a less score in the positive baggage, which are assigned a low weight. The final classifier classifies these samples as negative samples. Setting the detection threshold and training a complete classifier, we tend to acquire the required false positive rates and false negative rates. Retrain the initial weak classifier to get a zero false negative rate on the positive samples. Repeat the process to train the second classifier and yield a zero false negative rate on the remaining samples.

During the inference stage, given a testing image, we are able to treat every aspect in the Multiple Instance Learning model as a single class of human activity. Human activity identification requires a huge amount of training information; therefore it’ll lead in long training time. In this paper, we adopt an Iterative Querying Heuristic (IQH) algorithm to train the Multiple Instance Learning model.

The main step is as follows.

Input: Training baggage $\{x_1, x_2, \dots, M\}$, labels $\{y_1, \dots, y_m\}$, and parameters τ, w , and p , where τ is iterations times, p is instances per iteration, and w controls how many new instances, are considered in every iteration.

Step 1. Initialize $H_{ij}=0$, h_k being any classifier in $H(x)$.

Step 2. For $\tau = 1, \dots, \beta$.

Step 3. Query to new candidate instances per bag: $Z_i^r = I_i^{r-1} \cup \{P_i^i, \dots, P_\omega^i\}$, where $P_i^j \in x_i, \forall i$.

Step 4. Keep p the best rating instance by using h_r , where h_r is the kind of classifier and might label the instances of a positive bag x as varying with the latent parameter.

$$I_i^r \subset Z_i^r \text{ s.t. } |I_i^r| = \rho^- \text{ and } h_r^{r-1}(p) \geq h_r^{r-1}(p') \text{ for all } p \in I_i^r, p' \in Z_i^r \setminus I_i^r.$$

Step 5. Train h_r^{-r} with the opted instances:

$$h_r^{-r} \leftarrow \wedge (\{I_i^r, \dots, P_m^r\}, \{y_1, \dots, y_m\}).$$

Step 6. End for.

Step 7. Return h_r^β and the corresponding $h_r^{-\beta}$.

In Iterative Querying Heuristic, for positive baggage, we use the current estimate of the classifier to select the foremost positive instances, make sure that ensure a minimum of one of the queried instances is positive. For negative baggage, all instances are negative. In this case, we tend to choose the nearest instances to the decision boundary of our current classifier, which correspond to the toughest negative instances. Then the chosen instances are used to find a much better classifier. Thus, Iterative Querying Heuristic is expected to take advantage of a large amount of instances per baggage and it's not necessary to actually train with all of them at the same time. Each image has class label info in the training pictures that is very important for the classification task. This type of class label info is used to learn the Multiple Instance Learning model. So each image directly corresponds to certain human actions on training sets. Iterative Querying Heuristic algorithm makes the training more efficient and improves the entire recognition accuracy significantly.

5. Experimental Results and Analysis

In this section we presented some experiments to validate our proposed approach. The effectiveness of the proposed algorithm was verified by using C++ and Matlab hybrid implementation on a laptop with Pentium 3.2GHz processor and 4G RAM. We tested our algorithmic rule using Weizmann human action data set.



Figure 5: key frames for Weizmann human activity data set.

5.1. Experiment on Weizmann Data Set.

Weizmann human activity data set is the largest available video sequence dataset of human activities, whose sample pictures are shown in Figure 5. Weizmann database contains 83 video sequences and 9 groups of images, where each performs 9 different actions. They're "running," "walking," "jumping-jack" ("jack"), "jumping-forward-on-two-legs" ("jump"), "jumping-in place-on-two-legs" ("pjump"), "galloping-sideways" ("sideways"), "waving-two-hands" ("wave2"), "waving-one-hand" ("wave1"), and "bending," respectively. We track and stabilize the figures using the adaptive background subtraction masks that come with this data set. After an automatic preprocessing step, the video sequences were tracked and stabilized, and all the figures appeared in the center of the field of view. In this experiment, we studied recognition results of nine types of human activity from Weizmann human motion dataset. We investigated the performance of the proposed approach method for human action recognition. On the Weizmann dataset, the results of human activity recognition are shown in Table 1, where the correct recognition rate is defined as follows:

The correct recognition rate

$$\frac{\text{the number of times of correct recognition}}{\text{total number of the samples}} \times 100\%. \tag{19}$$

In Table 1 we can see that our method can correctly recognize most of human actions. The recognition rate is as high as 100% for "bending," "jumping-jack," "jumping-in-place on-two-legs," "galloping-sideways," "waving-two-hands," and "waving-one-hand". Our method achieves 98.5% average recognition rate. A few mistakes were confusions between "jumping-forward-on-two-legs" and "jumping-in-place-on two-legs" because these two kinds of actions were similar to each other.

Table1: The results of human activity on Weizmann database.

The action Categories	Total number of samples	The number of the times of correct recognition	The correct recognition rate (%)
Bending	200	200	100.0
Running	200	190	95.0
Walking	200	199	99.5
Jack	196	196	100.0
Jump	195	180	92.3
Pjump	192	192	100.0
Sideways	186	186	100.0
Wave2	192	192	100.0
Wave1	180	180	100.0
The average recognition rate			98.5

Conclusion

In this paper, we present a new method for human activity identification in video sequences, which focuses on two key problems extracting more useful and discriminating feature and improving the accuracy of classifier. The main contribution can conclude as follows. In feature extraction and representation, we extracted global silhouette feature using adaptive background subtraction method, and optical flow model was used for extracting motion features. Later on fusing these two types of features formed a hybrid feature vector. In activity modeling and recognition, we proposed the improved Multiple Instance Learning algorithm for human activity identification using Iterative Querying Heuristic (IQH) algorithmic rule, so that the recognition efficiency can be increased without compromising accuracy. Experiments were performed on Weizmann human activity data set Experiments evaluated the proposed method. Experimental results revealed that our proposed method performed better than previous ones. Our algorithmic rule can also recognize multiple actions in complex motion sequences.

References

- [1] A. Iosifidis, A. Tefas, N. Nikolaidis, and I. Pitas, "Multi-view human movement recognition based on fuzzy distances and linear discriminant analysis," *Computer Vision and Image Understanding*, vol. 116, no. 3, pp. 347–360, 2012.
- [2] H. Qian, Y. Mao, W. Xiang, and Z. Wang, "Recognition of human activities using SVM multi-class classifier," *Pattern Recognition Letters*, vol. 31, no. 2, pp. 100–111, 2010.
- [3] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions: a local SVM approach," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, vol. 3, pp. 32–36, IEEE, August 2004.
- [4] S. Danafar and N. Gheissari, "Action recognition for surveillance applications using optic flow and SVM," in *Computer Vision—ACCV 2007*, vol. 4844, pp. 457–466, Springer, Berlin, Germany, 2007.
- [5] Z. Lu, Y. Peng, and H. H. S. Ip, "Image categorization via robust LSA," *Pattern Recognition Letters*, vol. 31, no. 1, pp. 36–43, 2010.
- [6] M. B. David, Y. N. Andrew, and I. J. Michael, "Latent dirichlet allocation," *The Journal of Machine Learning Research*, vol. 3, pp.993–1022, 2003.
- [7] J. Shotton, T. Sharp, A. Kipman et al., "Real-time human pose recognition in parts from single depth images," *Communications of the ACM*, vol. 56, no. 1, pp. 116–124, 2013.
- [8] L. Xia, C.-C. Chen, and J. K. Aggarwal, "View invariant human action recognition using histograms of 3D joints," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '12)*, pp. 20–27, Providence, RI, USA, June 2012.
- [9] A. Jalal, M. Z. Uddin, J. T. Kim, and T.-S. Kim, "Recognition of human home activities via depth silhouettes and Rtransformation for smart homes," *Indoor and Built Environment*, vol. 21, no1, pp. 184–190, 2012.
- [10] B.Ni,G.Wang, and P.Moulin, "RGBD-HuDa Act: a color-depth video database for human daily activity recognition," in *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV' 2011)*, pp. 1147–1153, Springer, November 2011.
- [11] T.H.Thi, L. Cheng, J. Zhang, L.Wang, and S. Satoh, "Structured learning of local features for human action classification and localization," *Image and Vision Computing*, vol. 30, no. 1, pp. 1–14, 2012.
- [12] L. Ren, G. Shakhnarovich, J. K.Hodgins, H. Pfister, and P. Viola, "Learning silhouette features for control of human motion," *ACM Transactions on Graphics*, vol. 24, no. 4, pp. 1303–1331,2005.
- [13] M. Yamada, K. Ebihara, and J. Ohya, "A new robust real-time method for extracting human silhouettes from color images," in *Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 528–533, IEEE, Nara, Japan, April 1998.
- [14] J. M.McHugh, J. Konrad, V. Saligrama, and P.-M. Jodoin, "Foreground-adaptive background subtraction," *IEEE Signal Processing Letters*, vol. 16, no. 5, pp. 390–393, 2009.
- [15] A. Mokhber, C. Achard, and M. Milgram, "Recognition of human behavior by space-time silhouette characterization," *Pattern Recognition Letters*, vol. 29, no. 1, pp. 81–89, 2008.
- [16] S. Drgas and A. Dabrowski, "Kernel alignment maximization for speaker recognition based on high-level features," in *Proceedings of the 12th Annual Conference of the International Speech Communication Association (INTERSPEECH' 11)*, pp.489–492, August 2011.
- [17] M. J. Black and A. D. Jepson, "Estimating optical flow in segmented images using variable-order parametric models with local deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, pp. 972–986, 1996.
- [18] C.-I. Chang and A. Plaza, "A fast iterative algorithm for implementation of pixel purity index," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 63–67, 2006.
- [19] J. F. Murray, G. F. Hughes, and K. Kreutz-Delgado, "Machine learning methods for predicting failures in hard drives: a multiple-instance application," *Journal of Machine Learning Research*, vol. 6, pp. 783–816, 2005.
- [20] S. Ali and M. Shah, "Human action recognition in videos using kinematic features and multiple instance

learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no.2, pp. 288–303, 2010.

[21] J. D. Keeler, D. E. Rumelhart, and W. K. Leow, "Integrated segmentation and recognition of hand-printed numerals," in *Proceedings of the Conference on Advances in Neural Information Processing Systems 3 (NIPS '90)*, pp. 557–563, Morgan Kaufmann, San Francisco, Calif, USA, 1990.

[22] O. Maron and T. Lozano-Pérez, "A framework for multiple instance learning," in *Proceedings of the 11th Annual Conference on Neural Information Processing Systems (NIPS '98)*, pp. 570–576, December 1998.

[23] R. C. Bunescu and R. J. Mooney, "Multiple instance learning for sparse positive bags," in *Proceedings of the 24th International Conference on Machine Learning (ICML '07)*, pp. 105–112, ACM, Corvallis, Ore, USA, June 2007.

[24] S. Zhu, "Facial expression recognition based on MIL Boost," *Journal of Software*, vol. 9, no. 9, pp. 2435–2442, 2014.