Wireless image fuzzy recognition system for human activity

Hua-Ching Chen[1] · **Ching-Chang Wong**[2] ·
Hsuan-Ming Feng[1]

Abstract This study developed a fuzzy image model system for transmitting data over a wireless network channel to efficiently realize human activity in virtual images presentation. Because of the excellent mobile characteristics of wireless sensing networks, small devices are very desirable for local-area deployment. Complex model identification problems, such as acquiring and handling wireless image patterns, require analyzing a large amount of data, which occupies a long time at an acceptable transmission quality. In the proposed system, a cross-layer access method is employed to improve the visual clarity. Image packets are assigned to tune the category queue priority, with the probability allocated through a Markov chain model. This is a favorable approach to balancing the wireless image transmission traffic load. The similarity mixing algorithm, which is based on the maximal similarity and minimal disparity concepts, is used to aggregate the primary image features. The collected image patterns with converted coding vectors are efficiently trained through a human feature recognition procedure to generate a human model. A human action is received in real time from wireless sensing networks, and the image feature is retrieved by approximating a higher compatibility in practice simulations. The fuzzy image model uses the simple region-based evaluation and flexible extraction concepts to describe appropriate image partitions. This technology provides the highest possibility of human feature maps to identify the current action and offers a simple method for detecting human activity in indoor environments. Several human sensing and feature mapping experiments were conducted to verify the feasibility of applying the image recognition technology in nonlinear, time variant, and uncertain human activity problems. This study integrates numerous advantages from the mobility of wireless sensing; the proposed system efficiently controls congested image

✉ Hsuan-Ming Feng
hmfeng@nqu.edu.tw

1 Department of Computer Science and Information Engineering, National Quemoy University, No. 1 University Rd, Kin-Ning Vallage Kinmen, 892, Taiwan, Republic of China

2 Department of Electrical Engineering, Tamkang University, Tamsui District, New Taipei City, Taiwan, Republic of China

🖄 Springer

packages and easily confirms their related human activity. Experimental results verify that 60 testing frames approach about 96.6% accuracy within 3 s. These evaluations illustrate that it is applicable usage in some indoor environments.

1 Introduction

IEEE802.11 has become a widely used standard for wireless LAN (WLAN) applications, such as in factories, home care servicing, multimedia interaction, and traffic bottleneck monitoring. It is easily deployed with IEEE802.11 WLANs at enterprise locations, including airports, office buildings, campuses, training centers, stations, and home environments. Quality of service (QoS) issues of WLANs have been evaluated. In particular, short time delay and high transmission reliability have been addressed using the different multimedia data types of traffic services. The media access control layer plays a crucial role in video packet transmissions on the delivering channel of a wireless network. Several parameters, namely transmission opportunity (TXOP), arbitration interframe spacing (AIFS), and the contention window size (CW), are adjusted according to their related traffic-flow priorities in various service models. The system proposed in this article effectively allocates all resources on a wireless channel by improving wireless transmission rates to support image-based applications in indoor environments [3, 5, 15, 17, 20]. Most of the video content in wireless network mechanisms requires a large bandwidth to satisfy specific applications. This deficient condition restricts the high-speed performance of wireless systems. Most studies have explained that the original 802.11e mechanism selects specifics parameter values on the required multimedia channel to achieve high performance. Therefore, the adaption strategy presented in this paper to assign the multimedia flow on the accessing channel of the wireless network accelerates the transmission of video data.

The 802.11e standard offers more efficient QoS on wireless networks by using enhanced distributed channel access (EDCA) and hybrid coordination function controlled channel access (HCCA) to assign differential services [16, 19]. Research has shown that EDCA and HCCA improved the individual distributed and point coordination functions of 802.11, respectively. The fixed interframe space enters the waiting state during a time interval in which the wireless medium state is idle. Thus, the backoff timer procedure is initiated. Before the countdown step is completed, the traffic system restarts to deliver the image frames. In the 802.11e standard, four access categories (ACs), namely voice (AC_VO), video (AC_VI), best effort (AC_BE), and background traffic (AC_BK), are used to classify the package behavior in EDCA [12]. This study explored the EDCA mechanism as the main wireless video transmission method. The four ACs hold various transmission priorities to successfully complete their transmission roles. EDCA is a CW-based channel accessing machine that transmits streams with various ACs and guarantees the required bandwidth for high-priority applications by assigning different priorities on the basis of competitive access. Through the proposed control strategy, the transmission rate is improved using adaptive weights in the multimedia data layers. Wireless multimedia access parameters are adjusted to support the content of the specific characteristics in the broadcast image data. The hierarchical encoding process, which comprises restructuring of the lowest layer, contains the basic information for achieving the minimum basic perceived quality for images. The additional layer receives inherently unequal

priority by multimedia characterization to improve the overall image quality. Therefore, the perceptual cognition of the image coding stage by distinct priorities plays a crucial role in layered transmission and delay constraints to support perfect multimedia services. Video coding is an important issue for the efficient development of network transmission. The group of pictures (GOP) structure is an image concatenation technology that forms video streams. A GOP usually comprises the I-frame (intracode), P-frame (predictive code), and B-frame (bidirectional code). The goal of a GOP is to approach higher image quality by avoiding damaging encoded conditions. The I-frame contains the first delivery opportunity to obtain high priority for maintaining acceptable perceptual video quality. Primary medial data is more likely to be embedded in wireless transmission. However, multimedia transmission delay and unexpected bandwidth loss is inevitable when meeting the requirements of a wireless channel. Improvements in multimedia transmission with the 802.11e mechanism should be considered for various wireless applications. Wireless media access parameters should be adjusted to support detailed content with respect to the specific characteristics of image data. For video traffic delivery, the variety of video coding data must be understood.

In hierarchical encoding processes, the structure of the lowest layer contains the information to offer basic perceived image quality. An extra layer is proposed to improve the overall image quality. For example, the hierarchical coding technique is a novel method to improve the quality of video transmission [16, 19]. Such layer-based coding methods in image transmission play a primary control role in improving multimedia service on wireless networks. When the package delivery load is heavy, image information should be efficiently determined to appropriately allocate transmission space in other ACs. The crucial image information must be delivered via the highest priority queue. The present study's objective was to establish a comprehensive and effective image feature model for behavior recognition. The potential performance improvements of package control strategies are most obvious for high-throughput applications such as live video streaming and home care in indoor environments.

Image segmentation is a basic preprocessing technology for capturing salient features. In the proposed system, the shape characteristics of an image are obtained and these important features are gathered to describe the behavior within the picture. Numerous pattern learning techniques are used to collate human activity within a scene. Hongeng, Nevatia, and Bremond [11] developed a hierarchical process to present and model scenario events through trajectory shapes and features; every event is executed by a single actor. A real-time system was proposed by Amer, Dubois, and Mitiche [2] to detect context-independent events in single images. Assuming that objects were roughly segmented, they determined the low-level features, such as the shape and trajectory, of objects to infer four types of events. Ontology-based identification steps were provided by Fernández, Baiget, Roca, and González [9] to improve modeling of the most relevant events for human activity representation. A computer vision-based method developed by Gómez-Romero, Patricio, García, and Molina [10] focuses on tracking data, and their contextual information is generated from a symbolic model. A novel negative-silhouette-based method was used to extract human shapes and features in a recognition procedure. Region-based methods contain sufficient contour and texture information, which is relatively robust to noise and effectively recognizes human actions [13, 18, 21]. In common action methods, a model is built and postures are identified by tracking the changes of different limbs. Adaptable fuzzy clustering algorithms are used to reduce noise [8]. A fuzzy basis function classifier was constructed to recursively extract identified objects and identify activity [7]. A fuzzy rule-based classifier was developed using the recognition of the body posture and WiFi localization systems [1]. The fuzzy rules are automatically generated from a

dataset to generate a fuzzy model to achieve posture recognition. The constructed fuzzy system maintains a reasonable level of interpretability.

Wireless image fuzzy model architectures and human activity systems are described in the following sections.

2 Wireless image fuzzy recognition systems in human activity architecture design

Figure 1 shows a human feature extraction and action model. Details are discussed in the following two subsections.

2.1 Wireless image procedure

Figure 2 displays the architecture of a wireless network in an interior home environment. Each detected point is linked through an individual connection to the host server across the wireless network. In the home environment, an image model fuzzy platform is established using collections of several data streams from the distributed sensor in each subspace. The iBeacon is an ultra-small microprocessing computer with high volume and computational power. For iOS, Android, or Windows operating systems, Bluetooth low energy (BLE) technology is already a standard device in smartphones. Wireless networks are widely deployed in a large range of structures, and their ability to link reliably can vigorously promote more BLE applications. People carry smartphones with the iBeacon portable device. This smart device can easily detect whether people enter a specific indoor subspace. This condition initiates the camera to pursue the person's actions. A Web-based server receives the image streams of the
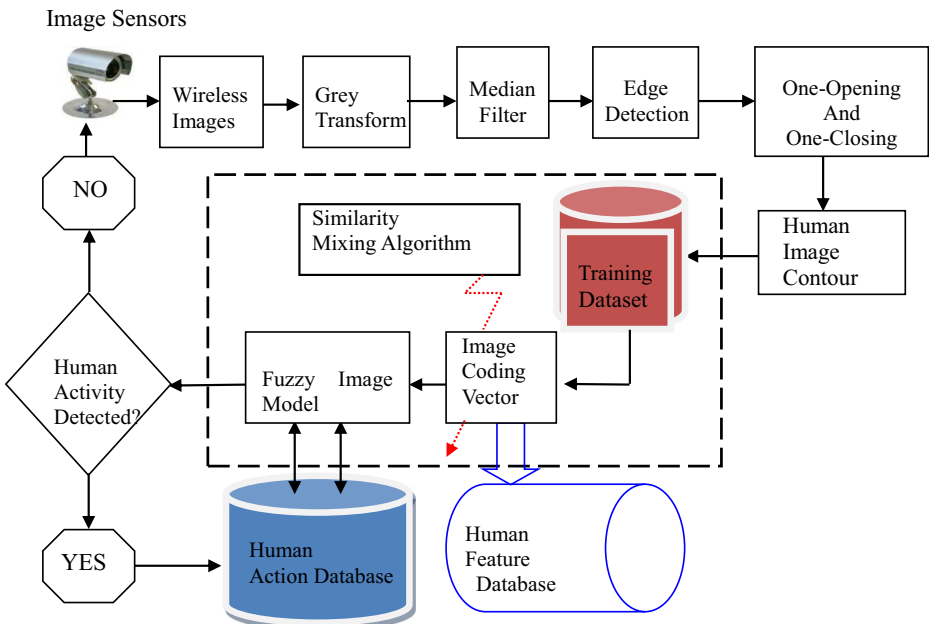

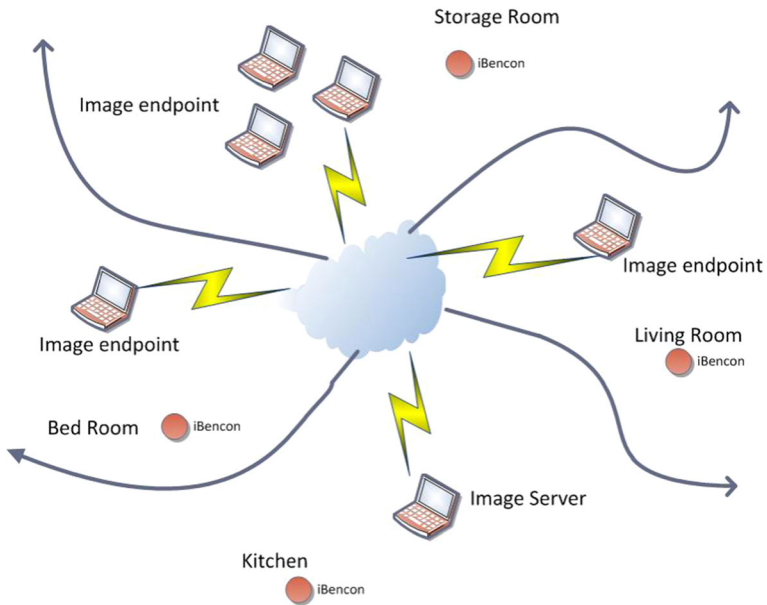**Fig. 1** Human feature extraction and action model

Fig. 2 Wireless network architecture in an interior home environment

body characteristics through the mobile transfer mode. An image database stores the specific feature and gradually traces the human model from the training database.

The MPEG4-based transmission mode is illustrated in Fig. 3. A collection of human images in video streams is successively analyzed to grade the GOP priority. A GOP is a basic unit of image sequences acquired through random access methods. This structure comprises the I-frame, P-frame, and B-frame. In the MPEG4 encoding process, the I-frame is the primary

Fig. 3 Wireless network image transmission mode
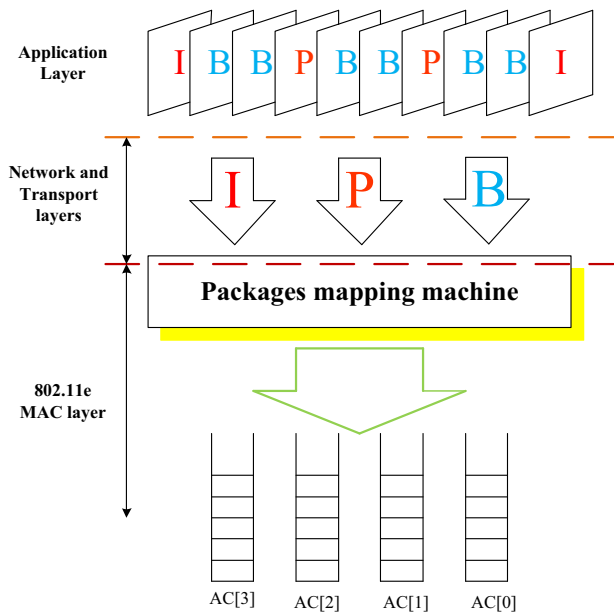
frame. The serial P-frame and B-frame refer to the I-frame information. An I-frame with substantial recognition properties requires protection on the wireless network transmission channel to avoid the loss of primary information and discard problems. The priority tuning strategy reduces package loss and damage to improve the QoS of image pattern transmission.

This research establishes an 802.11e EDCA wireless network image transmission model with Markov chain model concepts. In the Markov chain models, various internal image packages are transformed through the delivered media from the ending point to the working server. A time slot is constructed through mathematical inference with random and statistical operations. Because the time slots vary, data transmission and conflicting probabilities are computed according to the level of access priority to rebalance the data traffic flow. Data access evaluation indices, including throughput and time delay, are predicted based on the probabilities. Suppose that n points are located in a collection and each point's delivery grade is divided into N ($i = 0, 1, 2, \ldots, N-1$) ranks, where 0 denotes the highest priority. The term $b(i, t)$ represents the ith traffic flow class for backoff counters, and $s(i, t)$ is the back series number for the ith data flow class. According to [6], $\{b(i, t), s(i, t)\}$ is the ith image flow class to execute a 2D discrete Markov chain. Figure 4 shows the state transition diagram of the Markov chain model for the supposed ith AC (AC[i]).

In the time interval that includes package delivery procedures, the steady-state probability (τ_i) is the specific ith image data flow. p_i is the conditional probability of a conflict in the ith
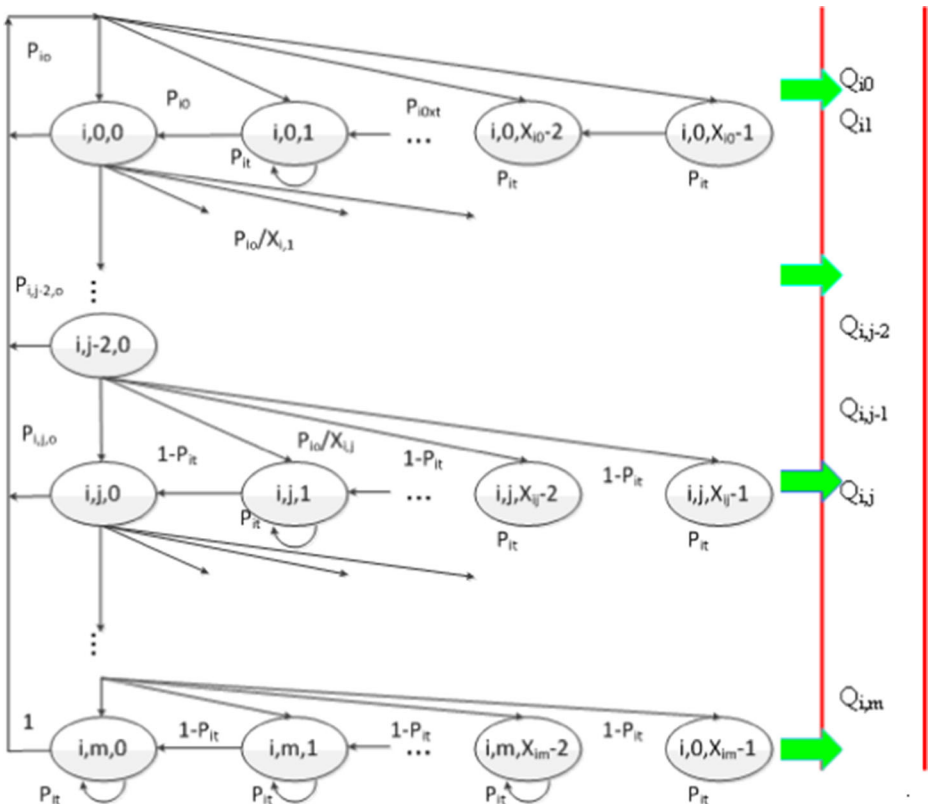

Fig. 4 Markov chain model for packages conveying states

image conversion traffic. The conflict includes real and virtual conflict conditions. The predicted probability of a traffic flow is illustrated as follows [6]:

$$Q_{ij} = \frac{1}{\sum\limits_{j=0}^{M} \frac{1-p_i}{1-p_i^{R+1}} P_i^j (1 + AE[b_j])} \tag{1}$$

where $AE[b_j]$ denotes the average backoff time, which is randomly selected for each ending site $[0, W_{i,j}]$. The competition window term $AW_{i,j}$ represents the ith class traffic in the jth backward progression. We define $AW_{i,0} = CW_{\min}[i]$. M is the maximum number of retransmissions, and m′ is the largest series. The exponential function reveals that $AW_{\max}[i] = 2^{m'} AW_{\min}[i]$. Therefore, the competition window size is created by

$$AW_{i,j} = \begin{cases} 2^j AW_{i,0}, & j \in [0, m'] \\ 2^{m'} AW_{i,0}, & j \in [m', m] \end{cases} \tag{2}$$

p_{it} is the probability term of detecting the busy channel for the ith class of traffic during the backoff cycle t. p_{io} represents the probability of conflict for the ith class of traffic sending an image stream when the backoff is 0. Then, the transition probability in the Markov chain expression is illustrated as follows:

$$\begin{cases} p\{i, j, k | i, j, k+1\} = 1 - p_{it}, & k \in [0, AW_{i,j} - 2], j \in [0, m] \\ p\{i, j, k | i, j, k\} = p_{it}, & k \in [0, AW_{i,j} - 1], j \in [0, m] \\ p\{i, j, k | i, j-1, 0\} = p_{io} / AW_{i,j}, & k \in [0, AW_{i,j} - 1], j \in [1, m] \\ p\{i, 0, k | i, j, 0\} = (1 - p_{io}) / AW_{i,0}, & k \in [0, AW_{i,0} - 1], j \in [0, m-1] \\ p\{i, 0, k | i, m, 0\} = 1 / AW_{i,0}, & k \in [0, AW_{i,0} - 1] \end{cases} \tag{3}$$

The proposed system uses hierarchical coding technology to extract the primary features of the trained image. Different image packets are layered according to their primary features to act as the controlled network flow indicators for approximating the required applications. In the image transfer process, more important packets with higher probability have higher priority to ensure superior image transmission services. An adaptive modulation obtains the capabilities of the current network load and realizes these situations to properly allocate image packets to approximate the most appropriate flow queue on a wireless network. An adaption mechanism regulates the variety of network traffic loads. This algorithm effectively controls congestion problems and provides optimal channel management to achieve higher transmission quality for wireless multimedia applications. The transmission probability is calculated as follows [6]:

(a) Determine a queue length:

$$QLEN(AC[s]) = (1 - AWi)QLEN(AC[s] - 1) + Wi * Qt(n) \tag{4}$$

(b) Compute the servicing probability type:

$$\text{Prob\_TYPE} = \begin{cases} 0; & QLEN(AC[s]) < THRESHOLD_{low} \\ 1; & QLEN(AC[s]) > THRESHOLD_{high} \\ \frac{QLEN(AC[s]) - THRESHOLD_{low}}{THRESHOLD_{high} - THRESHOLD_{low}} THRESHOLD_p; & \\ & THRESHOLD_{low} \leq QLEN(AC[s]) \leq THRESHOLD_{high} \end{cases} \tag{5}$$

The queue length of the access type in the individual range (i.e., s = 2, 1, 0) is defined as QLEN(AC[s]). The delivered queue length is distributed between the designed minimum

threshold (THRESHOLD_low) and maximum threshold (THRESHOLD_high). The maximum downward probability is denoted as THRESHOLD_p. The subsequent image transmission probability is calculated in the following formula:

$$Prob\_New = Prob\_TYPE * \frac{QLEN(AC[2]) - THRESHOLD\_low}{THRESHOLD\_high - THRESHOLD\_low} \tag{6}$$

2.2 Human feature extraction and action modeling procedure

Human actions are confirmed by each detected iBeacon when a person enters or departs the indoor environment. The collected image is transferred over the wireless network, and then image processing is applied to analyze image features and model the human motion in several indoor applications. The color image captured by a distinct photosensor generally comprises red, green, and blue (RGB) channels. Combining the acquired image with R, G, and B channel values is the traditional method for identifying the image shape. As shown in Eq. (7), the RGB color channels are first transformed into the Y, C_b, and C_r color space.

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.5 \\ 0.5 & -0.419 & -0.081 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{7}$$

The color contrast expressed in the Y C_b C_r channels is more productive than that in the RGB domain. Y represents light emission in the image and is regarded as the gray-level intensity. C_b and C_r are the related blue and red levels. Finally, the determinant of an individual pixel Y is assigned as the related gray level.

Images received through wireless network transformation are always expressed in the Y C_b C_r domain. Table 1 shows three image datasets, namely Stefan, hall_monitor, and coastguard, with their related packets and frame numbers (http://www.cipr.rpi.edu/resource/sequences/). For the designed wireless network topology and the layered package adaption method, Tables 2 and 3 show the frame loss number and package loss number with respect to the three image data sources for four video samples. Frame I has the highest priority, and the package and frame loss numbers are zero. Frame P contains the secondary important information and has an intermediate loss rate. Frame B is the least crucial frame. Simulation results illustrate that Frame B always has the highest loss number. Table 4 presents the average peak signal-to-noise ratio (PSNR) values for various video sources. Figure 5 shows

Table 1 Video frame and packet number settings in the video sources

Video	Format	Frame number			Total Number
		I	P	B	
Stefan	CIF	11	22	57	90
Hall Monitor	(352*288)	26	57	217	300
Coastguard		35	69	196	300
Video	Format	Packet number			Total Number
		I	P	B	
Stefan	CIF	29	15	36	80
Hall Monitor	(352*288)	70	41	137	248
Coastguard		94	49	125	268

Table 2 Frame loss number

Model	video	Stefan	Hall monitor	Coastguard
0		I = 0,P = 0,B = 12	I = 0,P = 0,B = 18	I = 0,P = 0,B = 21
1		I = 0,P = 1,B = 18	I = 0,P = 8,B = 28	I = 0,P = 11,B = 32
2		I = 0,P = 2,B = 24	I = 0,P = 12,B = 32	I = 0,P = 16,B = 36
3		I = 0,P = 4,B = 29	I = 0,P = 15,B = 35	I = 0,P = 19,B = 42

comparisons between original and wireless image results for selected video samples. These experimental results illustrate the high PSNR in the proposed wireless environment. Therefore, most of the transmitted images obtain a higher spatial resolution with an acceptable package loss number.

A popular Sobel edge scheme with arithmetical pixel-level computations is used to detect the borderline information. Edge detection is an essential technology to increase the degree of comparability to identify object boundaries. In practice, edge detection verifies edge lines through the partial comparison of pixel variations (http://www.cipr.rpi.edu/resource/sequences/) [4]. Two 3×3 masks in the form of an x-axle and y-axle are selected to execute Sobel operations. The masks are employed to evaluate the maximal pixel relation in the horizontal x-axis and vertical y-axis. Their derivatives for an evaluated image (EM) with the related x-axis and y-axis are determined by Eqs. (8) and (9), respectively [4, 14].

$$B_x = \left\{ EM_{(x+1,y-1)} + 2EM_{(x+1,y)} + EM_{(x+1,y+1)} \right\} - \left\{ EM_{(x-1,y-1)} + 2EM_{(x,y-1)} + EM_{(x+1,y-1)} \right\} \quad (8)$$

$$B_y = \left\{ EM+_{(x-1,y+1)} 2EM_{(x-1,y)} + EM_{(x-1,y-1)} \right\} - \left\{ EM_{(x-1,y-1)} + 2EM_{(x,y-1)} + EM_{(x+1,y)} \right\} \quad (9)$$

where the gradient magnitude for individual pixels in two points is calculated as $B = \sqrt{B_x^2 + B_y^2}$. If the threshold value $\gamma = 0.5$ is less than B, then B is assigned as an edge node. The Sobel operator removes insignificant pixels and intensifies the outstanding edge of objects. After the Sobel operation, the image appears to have a clean boundary on the region of interest. However, the Sobel edge operation also adds noise. Substantial noise causes lower quality in most image applications. Although the Sobel scheme is a simple and effective approach to finding the edges, it requires a valid noise filtering method. To reduce the noise influence and overcome image ambiguity, a center-weighted median filter (http://www.cipr.rpi.edu/resource/sequences/) [4] with a 3×3 window size is applied to suppress the noise and enhance the image area of interest. A standard median filtering machine is stimulated in the proposed system as follows:

$$MFILT_{(x,y)} = median\left\{ EM_{(x-k,y-1),(k,l)} \in w \right\} \quad (10)$$
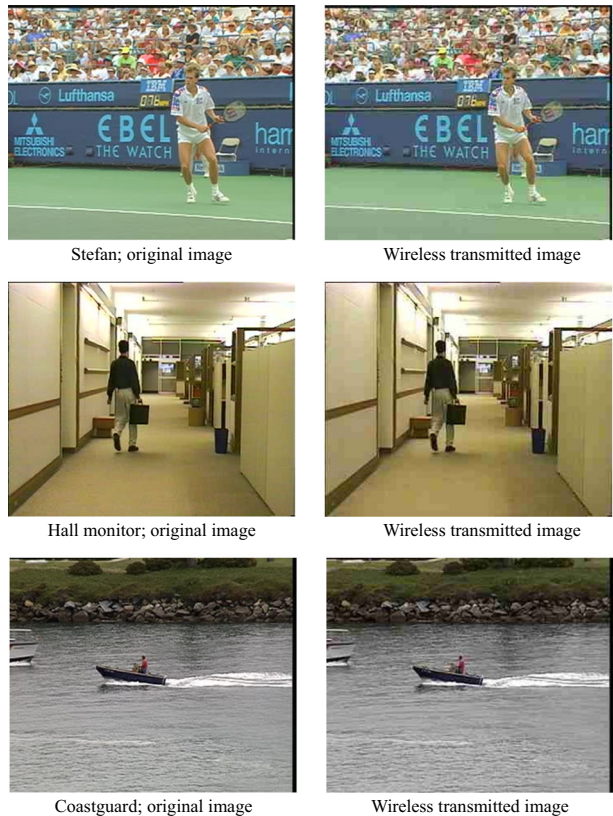
Table 3 Packet loss number

Model	video	Stefan	Hall monitor	Coastguard
0		I = 0,P = 0,B = 20	I = 0,P = 0,B = 28	I = 0,P = 0,B = 31
1		I = 0,P = 2,B = 37	I = 0,P = 15,B = 36	I = 0,P = 18,B = 38
2		I = 0,P = 3,B = 48	I = 0,P = 22,B = 44	I = 0,P = 25,B = 46
3		I = 0,P = 5,B = 60	I = 0,P = 25,B = 48	I = 0,P = 28,B = 50

Table 4 PSNR in three video sources

Model	Video	Stefan	Hall monitor	Coastguard
0		34.5568	34.2258	34.5266
1		33.8878	33.8978	33.9145
2		32.9009	32.8709	32.8786
3		32.0201	31.9887	31.9992

where *EM* denotes the gray-level value for the respective image pixel index in two coordinates (i.e., x-axis and y-axis). The selected value (k,l) is presented as the serial index in the form $(-1, 0, 1)$. Here, pixel (x, y) is the center position, and the median filtering machine regulates pixel values in a 3×3 neighborhood.

A human action training dataset (http://www.nada.kth.se/cvap/actions/) was employed to demonstrate the effectiveness of image extraction; Fig. 6 presents a human pose with various outlines. Figure 6a depicts an initial image received from the wireless network channel. Figure 6b shows the human pose after the completion of edge detection, without using the filtering machine. Some speckles appear in the output picture. Human posture features are detected and the region-based template technology extracts the required image information. The positions of four limbs and the object center are determined as derived features through the following steps: (Step 1) Find the distance between the center points of the object and endpoints of the limbs. (Step 2) Determine the angles

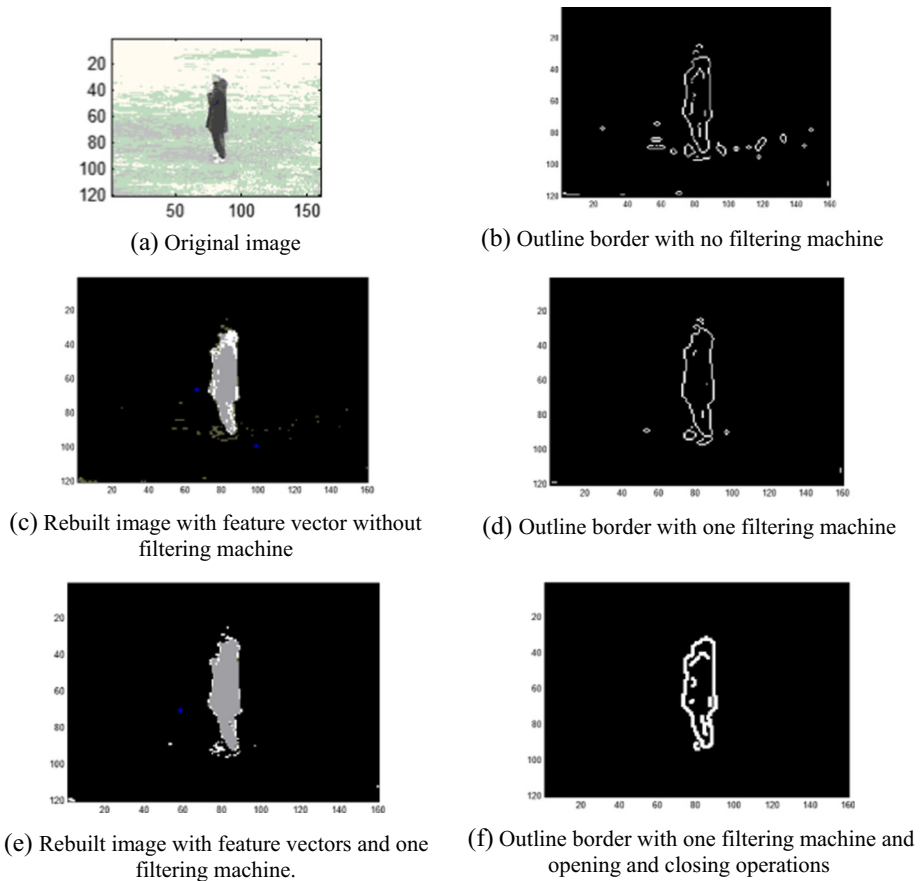**Fig. 5** Comparisons of original and wireless image results

Stefan; original image

Wireless transmitted image

Hall monitor; original image

Wireless transmitted image

Coastguard; original image

Wireless transmitted image

(a) Original image

(b) Outline border with no filtering machine

(c) Rebuilt image with feature vector without filtering machine

(d) Outline border with one filtering machine

(e) Rebuilt image with feature vectors and one filtering machine.

(f) Outline border with one filtering machine and opening and closing operations

Fig. 6 Human feature presentation

between the x-axis and the lines connecting the endpoints with the object center. (Step 3) Determine the angles between the y-axis and the lines connecting the endpoints with the object center. (Step 4) Each object creates features in a designed vector form. (Step 5) Features are stored in the database according to the vector dimensions. Figure 6c shows the image after application of the feature vectors without using the filtering machine. Figure 6d displays the outline border of a human pose with one filtering machine. The noise amplitude is successfully reduced by the image filter. The stage after elimination of the image noise in the detection procedure is shown in Fig. 6e. The image is rebuilt with feature vectors and one filtering machine to extract the possible object from the background. The visualized image is comprehensible, but fragments easily emerge from the mean of the border line of the object. One opening and one closing operation in image processing can fill the small gap when an unacceptable over-shape appears. The opening operation cuts the narrow connection and cleans small outliers. The closing operation smooths the narrow broken parts and collects thin gaps to repair small pieces. Figure 6f shows the rehabilitated image with a highlighted border after filtering machine, opening, and closing operations.

The similarity mixing algorithm with the max-min concepts is used to measure the similarity between the trained and targeted images. According to the max-min concepts, images with maximal degree of similarity are merged into the same groups, whereas data that have the minimal degree of similarity are separated into other regions. After completing the

computation for the training image dataset, the captured wireless image of body patterns is transformed into coding vectors, and the primary human features are extracted from the database. The fuzzy image model is generated according to the human action database, and it converts the input image coding vector and generates a feedback signal to retrieve this pattern's implication in real time. The human activity system includes human feature recognition and vector compatibility processes. In the human feature recognition process, a human feature is iteratively evaluated with max-min to obtain the trained coding vectors. The disparity between the captured and an estimated background images is determined according to their intensities. Wireless image subtraction of the region-based scenes is completed through perfect intensity comparison when the illumination is acceptable. That information is stored in the human feature database in the form of a coding vector to achieve the identified objective. In the vector compatibility processes, feature codes of wirelesses images are transformed and compared with current human coding vectors to match the desired objects and an applicable motion is chosen from the training patterns.

The fuzzy inference system receives the image coding vector in real time and accurately evaluates people's motion. The proposed image subtraction-based technology is a widely used, simple, and efficient method to detect moving objects.

The fuzzy inference system is applied to evaluate the human activity model in indoor environments. Input training data with n-dimensional variables ($x = (x_1, x_2, \ldots, x_n)$) and a hybrid-contoured membership function are considered as the premise of the fuzzy inference system. The membership function is derived as follows:

$$\mathrm{MR}_i\left(x\right) = \exp\left(-\eta\left(\sum_{k=1}^{n}\frac{\left(x-\alpha_k^i\right)^2}{\left(\omega_k^i\right)^2}\right)\right) \tag{11}$$

For the ith fuzzy rule with weight scale $\eta = 1$, α_k^i and ω_k^i affirm the center and length of the kth principal axis, respectively. The following fuzzy rule determines the desirable image similarity maps.

$$\text{Rule}^{(i)} \qquad : \qquad \text{IF } \underline{x} \text{ is } MR_i\left(x\right)_i$$

THEN $\underline{x}$ belongs to $y_{(j_1,j_2,\ldots,j_M)}$ with the condition of $CF_{(j_1,j_2,\ldots,j_n)}$

$$= 1, 2, \ldots, m, \text{ and } j \in \{1, 2, \ldots, M\} \tag{12}$$

Where m is the total number of fuzzy rules and M is the total training data number. $y_{(j_1,j_2,\ldots,j_M)}$ is the active consequent value constrained by the jth grade, $CF_{(j_1,j_2,\ldots,j_M)} > 0.5$.

In the proposed system, the weighted average defuzzifier is used to calculate the decision. Therefore, the output of the fuzzy inference system is obtained with the following formula:

$$Y_j = \frac{\left(\sum_{i=1}^{m}\mathrm{MR}_i\left(x\right)*y_i\right)_j}{\left(\sum_{i=1}^{m}\mathrm{MR}_i\left(x\right)\right)_j}. \tag{13}$$

The fuzzy image model efficiently approaches approximate similarity maps between the training and testing patterns.

# 3 Experimental results and analysis

In this study, several human images with different poses were employed to demonstrate the superiority of the proposed feature extraction method. Figure 6 exhibits image analysis for handclapping poses. Figure 7 and b displays the original image and its 3D graphical representation, respectively. In Fig. 7c, the received wireless image pattern after gray transform, median filter, and edge detection procedures is shown, and Fig. 7d displays the 3D meshed graphics of the received pattern. Figure 7e depicts the distributed segments of the handclapping image with the dimensional sizes of 4×7 pixels in the human feature recognition procedure. The similarity mixing
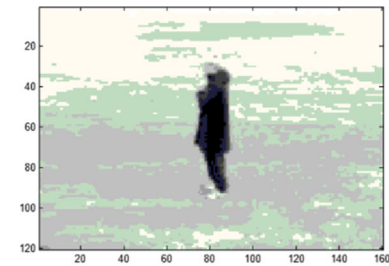
(a) Original handclapping image
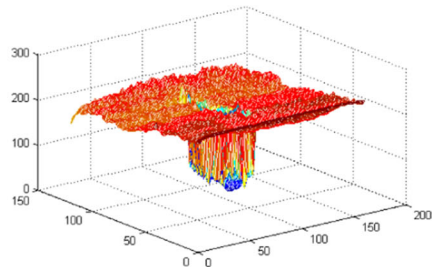
(b) 3D meshed image description for the original handclapping image
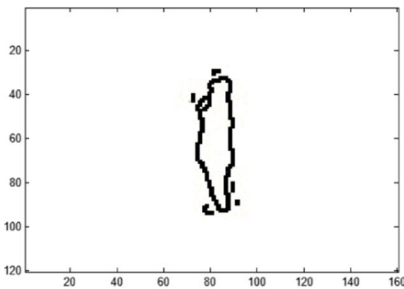
(c) Received wireless image pattern
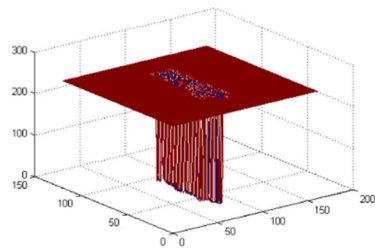
(d) 3D meshed image description for the related wireless image pattern
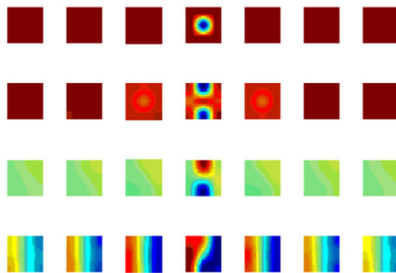
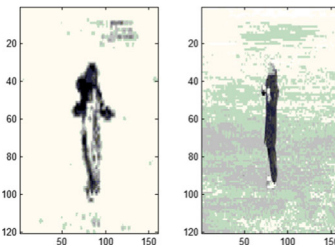(e) The distributed patterns of handclapping image with 4*7 sizes

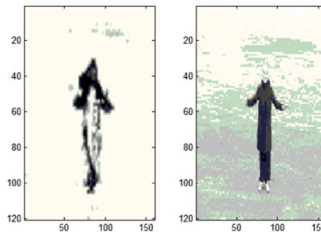**Fig. 7** Human action analysis in handclapping poses
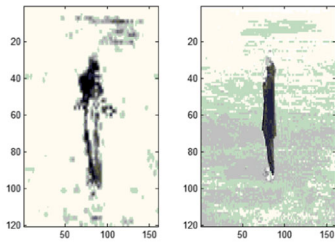
(f1) Similarity matching and rebuilding results and the related original handclapping pose in image 1
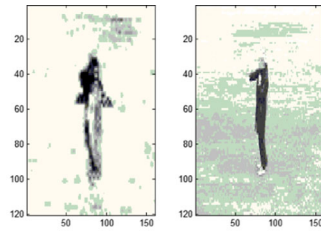
(f2) Similarity matching and rebuilding results and the related original handclapping pose in image 2

(f3) Similarity matching and rebuilding results and the related original handclapping pose in image 3

(f4) Similarity matching and rebuilding results and the related original handclapping pose in image 4

Fig. 7 (continued)

algorithm moves the maximal similarity data into the central region, and separates the minimal similarity data into the side areas. Four human handclapping action images were chosen to confirm the sampling plate. The most similar patterns contain similar pixel values and can be easily realized to rebuild the person's motion. Figure 7f1 shows a handclapping pose with two palms in front of the chest. Figure 7f2–f4 shows a handclapping posture with two palms spread at small, medium, and larges distances, respectively. The four similarity matching and building maps appear similar. The graphic model presents similar characteristics in contour, grain, and texture profiles. This study presents a successful model-based approach to identifying a person's motion through simple but efficient image processing.

In the second experiment, various hand waving images were examined. Figure 8a shows the original hand waving image. Its related 3D meshed description is illustrated in Fig. 8b. The image processing with gray transform, median filter, edge detection, and opening and closing procedures extracts the required human information after the transmission of wireless image patterns is complete (Fig. 8c). Figure 8d depicts the 3D meshed picture obtained through directly reference to the previous image pattern. To present the scattering situations of human hand waving motion through the proposed evaluation procedure, 4×7 pixel segmented regions are shown in Fig. 8e.
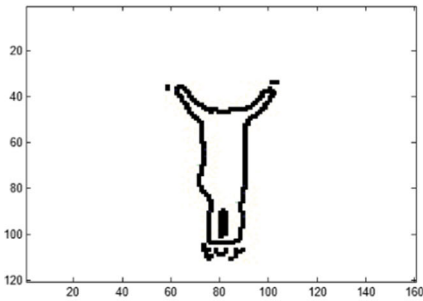
Four hand waving motion images with slightly different postures were employed to verify the correct motion through similar coding vector selection from the training feature maps.
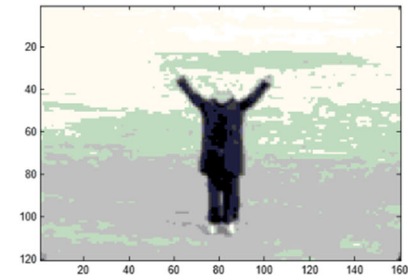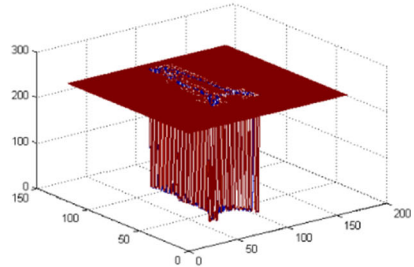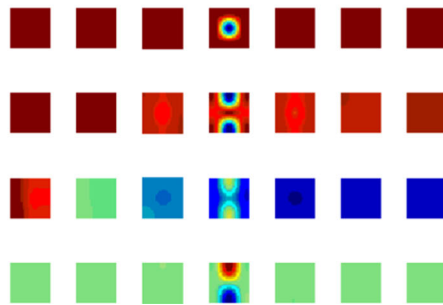
(a) Original hand waving image
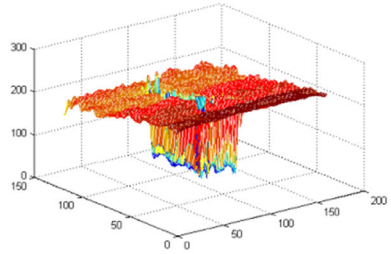
(b) 3D meshed image description for the original hand waving image

(c) Received wireless image pattern for hand waving

(d) 3D meshed image description for the received wireless image pattern

(e) The distributed patterns of hand waving image

Fig. 8 Human action analysis for hand waving poses

Figure 8f1 shows a hand waving posture with two arms down and near the chest. A posture with two arms spread is shown in the illustration of Fig. 8f2. In Fig. 8f3 and f4, the two arms are raised with large and small intersection angles, respectively. Human feature recognition through the similarity mixing algorithm executes the maximal similarity operation to approach the central region and separates the maximal variations to the outside area. The four matching and rebuilding pictures of people's motion appear similar. These similarities demonstrate that motion can be easily inspected. Images' contour, grain, and texture characteristics were also compared.
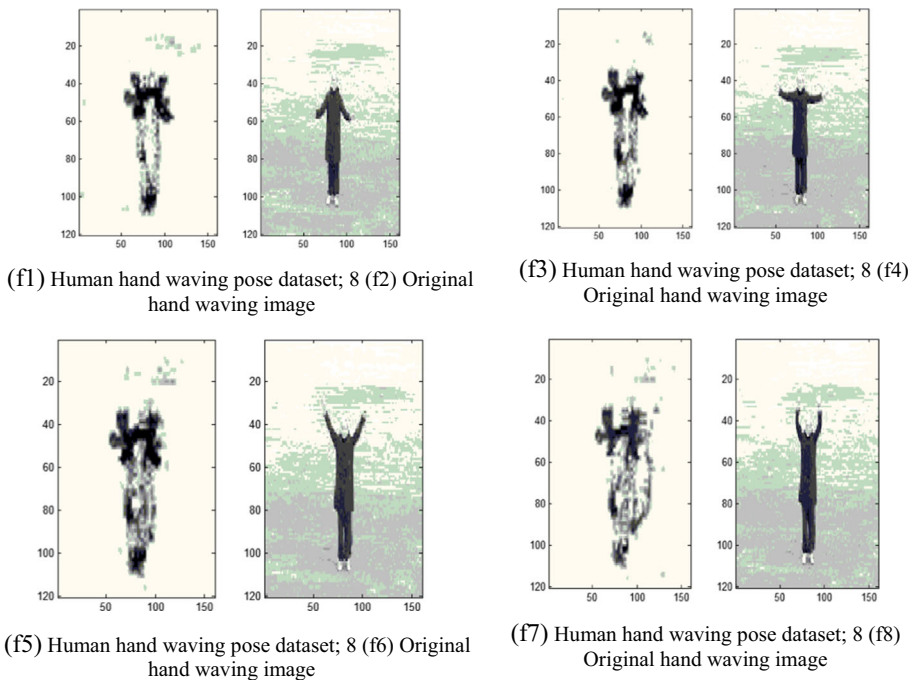
(f1) Human hand waving pose dataset; 8 (f2) Original hand waving image

(f3) Human hand waving pose dataset; 8 (f4) Original hand waving image

(f5) Human hand waving pose dataset; 8 (f6) Original hand waving image

(f7) Human hand waving pose dataset; 8 (f8) Original hand waving image

Fig. 8 (continued)

The last experiment involved examining the jogging posture shown in Fig. 9a, with distributed pixels in the form of a 3D meshed image illustrated in Fig. 9b. Figure 9c presents the received patterns from the wireless network channel and Fig. 9d shows the related gray pattern with the 3D meshed plots. Figure 9e shows 4 × 7 segmented patterns for the jogging posture.

A similarity mixing algorithm centralizes data with higher similarity into the same region and separates data with varying properties into the other areas. These patterns are compared with the most similar acquired feature to determine the correct evaluation of human motion. Figure 9f1 shows a jogging posture. The left leg slightly sags and two fists are at the side of the waist. Figure 9f2 shows a jogging state with two legs in parallel and two fists slightly outstripping the waist. Figure 9f3 shows a jogging posture in the opposite direction of that in Fig. 9f1. A jogging pose with the left leg slightly sagging and two fists not visible is displayed in Fig. 9f4. Because these mapping patterns are similar, a correct decision regarding the jogging posture can be made.

To evaluate the correctness in each human action, 60 different sample frames were extracted from the training dataset (http://www.nada.kth.se/cvap/actions/). These objects include hand clapping; hand waving and jogging actions by the similarity mixing algorithm are listed in Table 5. These results show that the Jogging behavior is reaching to 100%. Handclapping and Hand waving actions only fail one piece of image pattern. Total average accuracy rate is about 96.6%. It is obviously that the proposed method within 3 s can successfully match the desired pattern to recognize various human activities.
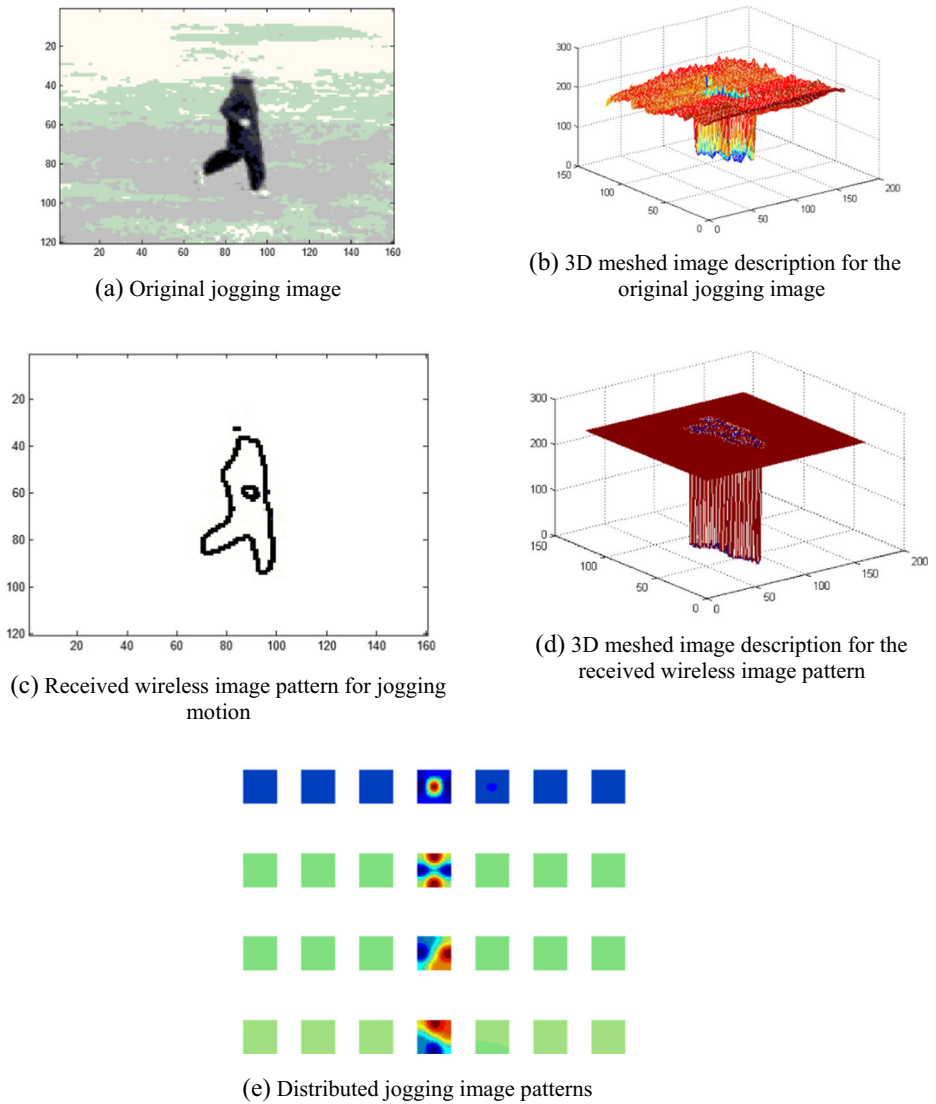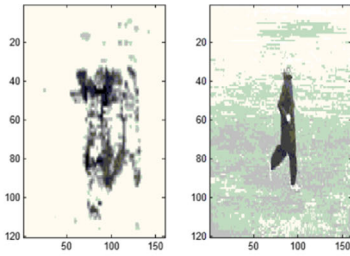
(a) Original jogging image

(b) 3D meshed image description for the original jogging image

(c) Received wireless image pattern for jogging motion

(d) 3D meshed image description for the received wireless image pattern

(e) Distributed jogging image patterns

Fig. 9 Human action analysis for jogging poses

4 Conclusions and future work

The system proposed in this paper contains a wireless image detection scheme using a layer mapping algorithm, human feature extraction, and active behavior machines. The system uses a representation of human features, known as the posture model, which contains all the information on a behavior.
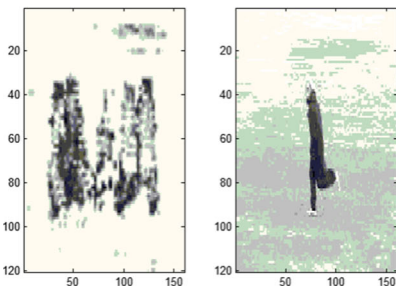
A layer adaption mapping algorithm with the predicted probability state calculated using a Markov chain model obtained the desired performance, verifying that the
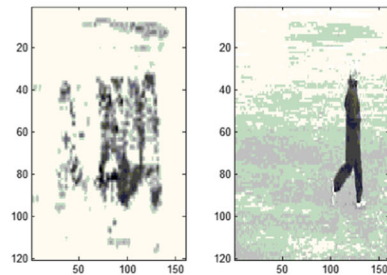
(f1) Similarity matching and building results and the related original jogging pose in image 1
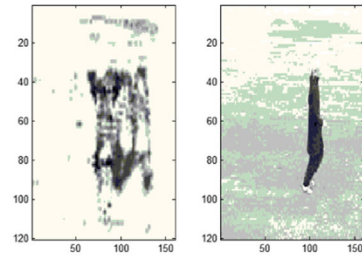
(f2) Similarity matching and building results and the related original jogging pose in image 2

(f3) Similarity matching and building results and the related original jogging pose in image 3

(f4) Similarity matching and building results and the related original jogging pose in image 4

Fig. 9 (continued)

video transfer quality was high. In experiments for three testing cases, high PSNR values were attained by dividing the data into three frames (I, P, and B) according to transfer priority.

Image processing, including gray transform, median filter, edge detection, and opening and closing procedures, enabled successfully rebuilding a distinct contour. For various image sequences, different human postures in the designed dimensional size patterns were rebuilt through a similarity mixing algorithm. The fuzzy inference system receives the temporal shape condition and determines the exhibited motion. Three experiments were conducted to analyze various motions. Region-based pattern recognition through the comparison of kernel vectors produced excellent similarity maps to evaluate human posture accurately in this study.

Table 5 Recognition data in three experiments

Patterns	Handclapping	Hand waving	Jogging	Accuracy
Handclapping	19	1	0	95%
Hand waving	1	19	0	95%
Jogging	0	0	20	95%
Average	X	X	X	100%

New smart home services are rapidly developing from conventional VoIP and video streaming to mobile gaming and interactive video gaming through natural human gestures or body motion. The other innovation applications, such as elderly care and monitor management system in home environment, physical fitness detecting system in indoor space and home-security system in intelligent building, could use multimedia delivery technologies based on the WLAN platform in the future.

References

1. Alberto AA, Alonso JM, Gracian T (2013) Human activity recognition in indoor environments by means of fusing information extracted from intensity of WiFi signal and accelerations. Inf Sci 233:162–182
2. Amer A, Dubois E, Mitiche A (2005) Rule-based real-time detection of context independent events in video shots. Real-Time Imaging 11(33):244–256
3. Chen CW, FENG Z (2013) Video over IEEE802.11 wireless LAN: a brief survey. China Commun 10(5):1–19
4. Chen HC, Feng HM, Lin TH, Chen CY, Zha YX (2016) Adapt DB-PSO patterns clustering algorithms and its applications in image segmentation. Multimedia Tools Appl 75(23):15327–15339
5. Chen HC, Feng HM, GUO D (2013) Novel MPEG-4 video cross-layer algorithm in wireless networks. Int J Comput Sci Eng 3(2):171–179
6. Chen HC, Guo DH, Feng HM (2014) Dynamic cross-layer adaptive mapping algorithm design for MPEG-4 video applications in wireless networks. Int J Innov Comput Inf Control (IJICIC) 10(1):1–12
7. Chen YP, Yang JY, Liou SN, Lee GY, Wang JS (2008) Online classifier construction algorithm for human activity detection using a tri-axial accelerometer. Appl Math Comput 205:849–860
8. D'Urso P, Massari R (2013) Fuzzy clustering of human activity patterns. Fuzzy Sets Syst 215:29–54
9. Fernández C, Baiget P, Roca FX, González J (2011) Determining the best suited semantic events for cognitive surveillance. Expert Syst Appl 38(4):4068–4079
10. Gómez-Romero J, Patricio MA, García J, Molina JM (2011) Ontology-based context representation and reasoning for object tracking and scene interpretation in video. Expert Syst Appl 38(6):7494–7510
11. Hongeng S, Nevatia R, Bremond F (2004) Video-based event recognition: activity representation and probabilistic recognition methods. Comput Vis Image Underst 96(2):129–162
12. IEEE Std 802.11e (2005) Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 8: medium access control (MAC) quality of service enhancements
13. Ignatov AD, Strijov VV (2016) Human activity recognition using quasiperiodic time series collected from a single tri-axial accelerometer. Multimedia Tools Appl 75(12):7257–7270
14. Jou SM, Chen CY, Feng HM (2013) Fuzzy surveillance systems through image extraction learning algorithm. ICIC Express Lett 7(3B):971–976
15. Kim JO, Tode H, Murakami K (2005) Service-based rate adaptation architecture for IEEE 802.11e QoS networks. In: IEEE global telecommunications conference, GLOBECOM'05, pp 3341–3345
16. Li XF, Ren T, Xu J (2010) A cross-layer design for transmission of scalable H.264 video over IEEE 802.11e networks. International Conference on Computational Problem-Solving (ICCP), pp 306–309.
17. Majkowski J, Palacio FC (2006) Enhanced TXOP scheme for efficiency improvement of WLAN IEEE 802.11e. 2006 I.E. 64th In Vehicular technology conference, VTC-2006 Fall, pp 1–5
18. Rahman SA, Leung MKH, Cho SY, Human action (2013) recognition employing negative space features. J Vis Commun Image 24:217–231
19. Ramos N, Panigrahi D, Dey S (2007) Dynamic adaptation policies to improve quality of service of real-time multimedia applications in IEEE 802.11e WLAN networks. Wirel Netw 13(4):511–535
20. Seyed VA, Ozgur G, Ozgur E (2016) QoS based aggregation in high speed IEEE802.11 wireless networks. 2016 Mediterranean Ad Hoc networking workshop (Med-Hoc-Net), pp 1–7
21. Wang L, Hu W, Tan T (2003) Recent developments in human motion analysis. Pattern Recogn 36:585–601

Hua-Ching Chen received the B.S. and M.S. degree in business administration and Computer Science and Information Engineering from National Quemoy University, Kinmen, Taiwan, R.O.C., in 2008 and 2010. He received Ph.D. degrees in Electronic Engineering from Xiamen University, Xiamen, Fujian, China, P.R.C., in 2014, respectively. He is currently the lecturer with the Department of Computer Science and Information Engineering, National Quemoy University. His current research interests include wireless networks, cloud computing, fuzzy systems, optimal learning algorithms, image processing and robot system.

Ching-Chang Wong received the B.S. degree in electronic engineering from Tamkang University, Taipei, Taiwan, in 1984, and the M.S. and Ph.D. degrees in electrical engineering from Tatung Institute of Technology, Taipei, in 1986 and 1989, respectively. In 1989, he joined the Department of Electrical Engineering, Tamkang University, where he is currently a Professor. His research interests include fuzzy systems, intelligent control, system-on-a-programmable-chip design, and robot design.

Hsuan-Ming Feng received the B.S. degree in automatic control engineering from Feng-Chia University, Taichung, Taiwan, R.O.C., in 1992. He received M.S. and Ph.D. degrees in Computer Science and Information Engineering from Tamkang University, Tamsui, Taipei Hsien, Taiwan, R.O.C., in 1994 and 2000, respectively. He is currently the full professor with the Department of Computer Science and Information Engineering, National Quemoy University. His current research interests include fuzzy systems, neural networks, wireless networks, optimal learning algorithms, image processing and robot system.