sustainability

MDPI

Article

Infrared Human Posture Recognition Method for Monitoring in Smart Homes Based on Hidden Markov Model

Xingquan Cai *, Yufeng Gao, Mengxuan Li and Wei Song *

School of Computer Science, North China University of Technology, Beijing 100144, China; g17gyf@sina.com (Y.G.); mengxuanli_amy@126.com (M.L.)
* Correspondence: caixingquan@ncut.edu.cn (X.C.); sw@ncut.edu.cn (W.S.); Tel.: +86-10-8880-1550 (X.C.); +86-10-8880-1991 (W.S.)

Academic Editors: James Park and Marc Rosen
Received: 31 May 2016; Accepted: 12 August 2016; Published: 3 September 2016

Abstract: Smart homes are the most important sustainability technology of our future. In smart homes, intelligent monitoring is an important component. However, there is currently no effective method for human posture detection for monitoring in smart homes. So, in this paper, we provide an infrared human posture recognition method for monitoring in sustainable smart homes based on a Hidden Markov Model (HMM). We also trained the model parameters. Our model can be used to effectively classify human postures. Compared with the traditional HMM, this paper puts forward a method to solve the problem of human posture recognition. This paper tries to establish a model of training data according to the characteristics of human postures. Accordingly, this complex problem can be decomposed. Thereby, it can reduce computational complexity. In practical applications, it can improve system performance. Through experimentation in a real environment, the model can identify the different body movement postures by observing the human posture sequence, matching identification and classification process. The results show that the proposed method is feasible and effective for human posture recognition. In addition, for human movement target detection, this paper puts forward a human movement target detection method based on a Gaussian mixture model. For human object contour extraction, this paper puts forward a human object contour extraction method based on the Sobel edge detection operator. Here, we have presented an experiment for human posture recognition, and have also examined our cloud-based monitoring system for elderly people using our method. We have used our method in our actual projects, and the experimental results show that our method is feasible and effective.

Keywords: human-computer interaction; feature extraction; Hidden Markov Model; human action recognition

---

1. Introduction

As we all know, the smart house is a platform that integrates network communication technology, intelligent electronic technology, safety technology, remote control, lighting control, surveillance, etc. Smart homes are based on an efficient and environmentally friendly family affairs management system, improving the quality and promoting the harmony of household life. Smart homes are the most important sustainability technology of our future. In smart homes, intelligent monitoring is an important component, because monitoring systems can detect the changes in the environment including that of air quality, temperature, and humidity, and can detect human behavior. Today, elderly people face the threat of sudden illness, such as heart disease and cerebral vascular disease, which is especially problematic when they live alone. So, there is a need for a human posture recognition system for monitoring in the smart homes of elderly people. In this system, we can recognize sudden illness in old people automatically and contact the hospital.

In recent years, with the rapid development of virtual reality, augmented reality, and interactive multimedia, human-computer interaction (HCI) technology is in higher demand than ever before [1]. In particular, human posture recognition can be used in monitoring in smart homes, interactive entertainment, and in the identification of criminal behavior. However, there is no effective method for human posture detection. So, in this paper, we provide an infrared human posture recognition method for monitoring in sustainable smart homes based on a Hidden Markov Model (HMM).

Currently, the management technology for memory and storage is very advanced and the speed of internet and mobile networks is very fast. Also, wide angle cameras in smart homes are very sensitive when taking pictures. So, we need to establish a human posture recognition method based on cloud-computing, and develop a sustainable human posture recognition system to monitor abnormal behavior in different people to avoid unnecessary risks.

In order to optimize the interaction between the human and the computer, we need to let the computer understand the human's commands [2]. The computer needs to understand the different postures and movements of humans [3]. Human posture recognition technology can identify the user's postures, and convert these postures into information [4]. Then, it compares the actual behavior with the default behavior. According to its maximum similarity, we can then classify it. Scientists have conducted this research based on motion detection and feature extraction. Based on the analysis of human movement patterns, the relationship between video content and action type has been established [5]. The interaction between human and computer is realized through this information. It can achieve the ultimate goal of making computers "look at" and "understand" videos. Therefore, using a computer to recognize the postures of the human body has been a hot topic in recent years and remains so today.

This paper is written in accordance with the principles of Future Sustainability Computing (FSC) [6]. Through innovation and optimization of past methods, this paper realizes a method to support FSC. Athletic human posture recognition is one kind of human posture recognition, concerning pattern recognition. It will match the obtained sequence to the typical sequence, which is obtained by pre-calibration in order to classify the test sequences [7,8].

At present, the main method for human posture recognition is the probability network method [9,10]. There are two ways of using the probability network method in the human network, which are the Bayesian network and the HMM [11–13]. Methodically, this paper mainly focuses on human posture recognition based on the HMM. In addition, the key technologies of human-computer interaction are also studied. For the human movement target detection, this paper proposes a method based on a Gaussian mixture model [14]. Furthermore, this paper proposes a method based on the Sobel edge detection operator for the extraction of the human's object contour [15].

2. Background

Currently, the main human posture recognition methods are the template matching method and the probability network method [16,17]. Among them, the template matching method has been used earlier. Today, however, the probability network method is more widely used in human posture recognition. The template matching method transforms the image sequence into a set of static templates and then matches the template of the test sequence to a defined reference sequence. In this way, it can obtain a recognition result. This method is simple and has lower computational complexity, but it is sensitive to motion interval change and noise. The probability network method defines each static body posture as a state and changes the various states by calculating the transition probability. Every posture sequence is an all-state traversal. When the recognition of a human posture is made, it first calculates the joint probability of each traverse. After that, the sequence can be classified according to the obtained joint probability. The probability network method has high computational complexity.

However, it has a better stability to deal with an action sequence's small changes in both time and space. This is the reason for its current and frequent use.

In addition, there are the following successful methods for posture recognition: In 2012, Zhao and Li [18] presented a recognition algorithm of human posture based on multi-feature fusion. Through background subtraction and shadow elimination, they obtained human motion silhouettes and contours. Thus, by describing the transformation of human silhouette features, they obtained a new feature that had higher discriminating power by Karhunen-Loeve (K-L) transform being a transformation based on the statistical characteristics [19]. However, their model is problematic, as it is easy to make mistakes.

In 2012, Wu, Deng, and Kang [20] presented a method to express human action in complex scenes with local feature description. In their method, a scale-invariant key point detector and a 3D-Harris detector were used to find the local key points in video samples. The existing local feature descriptor and shape descriptor are employed to describe the structural information about the positions of the key points. After that, the bag-of-features model is utilized to calculate the distribution of the features. Finally, the fuzzy integral scheme is used to fuse the local features.

In 2014, Chen [21] presented a novel method for action recognition by combining a Bayesian network model with a high-level semantic concept. They extracted spatial-temporal interest points and 3D-SIFT descriptors around each interest point in the videos. Then, low-level features were used to describe these videos before building the projection from low-level features to high-level features through training attribute classifiers. Finally, a behavior recognition model based on an attribute Bayesian network was established.

In 2015, Liang [22] presented an action recognition method on the motion process of a human's lower limbs. According to the motion characteristics of the human's lower limbs, motion parameters used for action recognition were chosen. Then, they presented an action characteristic extraction method based on wavelet fractal and the least squares fit method. After that, the action recognition of each motion parameter of the human's lower limb kinematics chain was realized based on a support vector machine. Finally, on the basis of the action recognition result of each motion parameter, evidence theory was employed to fuse the recognition results. However, with the increase of the action types and the training data, the training time of Support Vector Machine (SVM) [23] increases greatly and the accuracy rate is reduced under certain training parameters.

So, in this paper, we focused on developing a method for human posture recognition based on the HMM [24,25]. We also provide a human movement target detection method based on the Gaussian mixture model [14] and a human object contour extraction method based on the Sobel edge detection operator [15].

3. Method of Human Posture Recognition Based on the Hidden Markov Model (HMM)

3.1. The Basic Principle of the HMM

The Markov process can be defined as when the next state of a stochastic process is only related to the previous n states, and has no relation with states before the previous n states. Then, the stochastic process is called an n-order Markov process. The HMM is a 5-tuple, and a standard HMM can be represented as in Equation (1):

$$\lambda = (N, M, \pi, A, B) \tag{1}$$

The parameters are expressed as follows: The variable N is the number of hidden states in the model. It is denoted as $(\theta_1, \theta_2, ..., \theta_N)$. When we set the state of the time t as q_t, there are $q_t \in (\theta_1, \theta_2, ..., \theta_N)$. The variable M refers to the corresponding number of observation states of each hidden state in the HMM. The observation state corresponds to the output value of the model, which is the observation state in the state of $(V_1, V_2, ..., V_M)$. If the observation state at time t is O_t, then $O_t \in (V_1, V_2, ..., V_M)$.

Matrix A represents the transfer-probability matrix of state $A = \left(a_{ij}\right)_{N\times N}$ with a_{ij}, as shown in Equation (2):

$$\alpha_{ij} = P\left(q_{t+1} = \theta_j | q_t = \theta_i\right), 1 \leq i, j \leq N \tag{2}$$

where α_{ij} indicates the probability of state transfer from state i to state j, which satisfies $a_{ij} \geq 0, \forall i, j$, and $\sum\limits_j a_{ij} = 1, \forall i$.

The matrix B indicates the probability distribution matrix of the observation value $B = \left(b_{jk}\right)_{N\times M}$, where b_{jk} is defined as in Equation (3):

$$b_{jk} = P\left(O_i = V_k | q_i = \theta_j\right), 1 < i < N, 1 < k < N \tag{3}$$

where the variable b_{jk} represents the probability of generating the corresponding observed value k in the state j.

The variable II represents the probability vector of the initial state $\pi = (\pi_1, \pi_2, ..., \pi_N)$, where π_i is shown in Equation (4):

$$\pi_i = P\left(q_1 = \theta_i\right), 1 < i < N \tag{4}$$

That is to say that if the probability of state θ_t at time $t = 1$, π_i satisfies $\sum\limits_i \pi_i = 1$.

3.2. Establish a HMM for the Movement of the Human Body

This chapter will classify the recognition of human behavior as time-varying data. In the process of classification, every human posture will be defined as a state at first. The link between the various states is made by probability. The posture that the user has in the scene is called a series of observations. Every posture of the user can compose a motion sequence. Each motion sequence is needed to traverse the states defined before. In this process, we calculate the joint probability and set its maximum value as a criterion for action recognition to classify the human action.

In this paper, the human posture recognition method establishes the relationship between adjacent states with the HMM. Accordingly, we assume that the observation sequence is decided by a hidden process, which is composed of a fixed number of hidden states. That is a random state mechanism.

The model parameter $\lambda = (N, M, \pi, A, B)$ of human posture recognition is described in Section 3.1. N represents the number of behavioral categories, and M is the number of key gestures of a basic behavior database. That is the size of the code. This paper uses four key gestures to describe each behavior. Thus, $M = 4N$.

3.3. Model Parameter Training

HMM parameter training plays an important role in the process of postures recognition. It is also called parameter estimation. Whether the parameter estimation is reasonable directly affects the accuracy of the postures recognition results. Therefore, before the posture recognition, the known postures sequence must be defined to train the parameter $\lambda = (\pi, A, B)$ set.

The training process of the model parameters has includes the following steps:

At the beginning, set the initial parameters. This paper manually initializes the model parameters based on prior knowledge. Then, train the parameters reasonably and effectively. In this paper, the estimation of model parameters is completed by the Baum-Welch algorithm. Firstly, determine three initial parameters according to the initial model λ_0. They are the initial probability distribution π, the initial transfer-probability matrix A, and the confusion matrix B. Secondly, use the multiple observation samples sequence revaluation equation of the Baum-Welch algorithm and a multiple iterative algorithm to train a new mode, λ. Use O to represent the observation sequence. Third, use the forward algorithm to calculate the output probability. If the value of $\log P\left(O|\lambda\right) - \log P\left(O|\lambda_0\right)$ is less than the value of Δ, the algorithm process is completed. Otherwise, reset the parameter to make the $\lambda_0 = \lambda$, return to the second step, and repeat the algorithm.

3.4. Target Motion Human Posture Recognition

The decision of a target human posture recognition method is decided by the bias' maximum posterior (MAP) estimate. Input the unknown observed sequence X_0 into the model to calculate the log likelihood probability value. Obtain the state of the maximum output probability. At this time, the posture most likely belongs to a behavior of the unknown behavior categories. The process is as follows:

Make an unknown human posture sequence $X_t = \{O_1, O_2, ..., O_T\}$. The parameter T is the length of the observation sequence. The unknown behavior O_t generates a likelihood probability of λ. Then, use the forward algorithm to calculate the probability $P\left(O_c|\lambda\right)$. The variable c_t represents the observed value of an observation sequence in time t. i and j represent each state. The variable π represents the initial probability of state i. The function $b_t\left(c_t\right)$ represents the probability that state i generates an observation value c_t in time t. The function $a_t\left(j\right)$ represents the transition probability from the current state i to the next state j. The likelihood probability $P\left(O_t|\lambda\right)$ of behavior O_t and model λ are calculated by $\sum\limits_{i=1}^{N} a_T\left(i\right)$; here, $\left(O_t = \{c_1, c_2, ..., c_T\}\right)$. As mentioned above, according to the behavior model parameter λ, the behavior which has the maximum likelihood probability is identified as the target behavior in the unknown behavior sequence.

4. Human Movement Target Detection Method Based on the Gaussian Mixture Model

In order to improve the real-time detection and reliability of the moving target detection, this paper describes a moving target detection module. This module is established with consideration of scene and illumination changes. The according background subtraction method is based on the Gaussian mixture model. It can detect the human movement target of the scene.

The specific process is as follows: Firstly, the infrared camera collects the video image of the scene and transmits the video image to the computer. The computer processes the received video image to reduce the effect of the noise. Secondly, the Gaussian mixture model is established on the basis of pretreatment. This model can determine the region of the moving target. Thirdly, the video image is processed by erosion and dilation: In morphology, the purpose of erosion is to eliminate object boundary points. The effect of erosion algorithms in this paper is to eliminate isolated foreground points. By contrast, the dilation operation aims to expand object boundary points. In this paper, dilation is employed to fill the foreground. Finally, in this manner, the target of the human movement is detected.

4.1. Pretreatment

The quality of the video images may be reduced in the process of generation and transmission. The reason is the image in these processes will be affected by a variety of noises and these noises will also affect the post-treatment. Causes of noise can be the camera noise, the instantaneous ambient noise of the natural environment in the scene, and other noises. In the process of image operation, image de-noising is a very important step. The de-noising operation can reduce the impact of the processing results; improving the quality of the image and the ease of posting the process images. The purpose of the filtering pretreatment before moving target detection is to reduce the noise points and to prevent the error of recognizing noise points as moving targets. The de-noising operation can improve the accuracy of detection and achieve the purpose of facilitating subsequent processing at the same time.

At present, there are two kinds of filtering methods: the spatial domain method and the frequency domain method. The median filtering algorithm is simple and easy to implement. At the same time, the median filtering algorithm can also smooth images, enhance edges, and save information. Therefore, this paper uses the median filtering method for pretreatment.

In two dimensions, the median filtering algorithm is defined as follows: Let $\{x_{ij}\}$ represent the gray value of various points in the image. For filter window A, its size is expressed as

$N = (2K+1) \times (2K+1)$, and y_{1j} is the median of window A in x_{1j}. The corresponding equation is as follows:

$$y_{1j} = med\left\{x_{1+r,j+s}\ , (r,s) \in A\right\} \tag{5}$$

4.2. Establish a Gaussian Mixture Background Model

The core of the background subtraction method is to establish an effective and reliable background model. The focus of establishing the background model is to make the model sensitive to moving targets and to make it have enough robustness when the background environment changes. The background model is the representation of the scene. Therefore, the complexity of the background model is determined by the complexity and changes of the scene. In this paper, the system scene is very complex, so we propose a background module method based on the Gaussian mixture background model.

The basic principle of the single Gaussian model is as follows: Establish the Gaussian function model for every pixel in the scene. At this time, each pixel automatically obeys the distribution of the average μ and the standard deviation σ. The Gaussian distribution of every point is independent. Each pixel processes a collection, which is a series of point values in the corresponding period. This means that at any time t, the value of the pixel (x,y) should be one point of a point set which is on the axis of time, as shown in Equation (6):

$$\{\mathrm{X}_1, X_{2,\ldots,} X_t\} = \{I(x,y,i) | 1 \le i \le t\} \tag{6}$$

The variable I represents the image sequence, and i represents the frame number, i.e., the i-th frame. The collection of these points is in accordance with the Gaussian distribution, as shown in Equation (7):

$$P\left(X_t\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X-\mu)^2}{2\sigma^2}} \quad X \in I\left(x,y,t\right) \tag{7}$$

Most real scenes are in motion and often have a few small changes. These small changes enable the variation of the gray value of the same background point at different times. In this case, employing a single Gaussian model cannot describe the background image accurately. Therefore, this paper uses the Gaussian mixture model to represent the background image.

The Gaussian mixture model is based on a single Gaussian model and establishes the model of each pixel point according to the logic of multiple Gaussian models. The basic idea is: First, represent the characteristic of each pixel using a Gaussian mixture model. Second, after each time period, the current background is represented by a subset of the Gaussian mixture model. Third, update the Gaussian mixed model with each new image frame. Finally, if the current image pixel matches the Gaussian mixed background model, then that point will be judged as the background. Otherwise, determine a point as the foreground point. We also get good experimental results if the Gaussian mixture model is used to detect the moving target in a more complex environment.

In the sequence of video images, we determine the gray values for any pixel $\{x,y\}$ from moment 1 to moment t. We express the gray values as $\{I_1, I_2, \ldots, I_t\}$. The probability density function of the K-th Gaussian distribution at time t is shown in Equation (8):

$$f\left(I_t \middle| \mu_{k,t}\right) = \frac{1}{(2\pi)^{\frac{n}{2}} \left|\sum_{k,t}\right|^{\frac{1}{2}}} e^{-\frac{1}{2}(I_t-\mu_{k,t})^T \sum_{k,t} (I_t-\mu_{k,t})} \tag{8}$$

Among them, $\mu_{k,t}$ expresses the average, and $\sum_{k,t}$ expresses the covariance matrix.

The current characteristics of pixel K are related to the characteristics of pixels at a past moment in time. The probability of pixel $I\,(x, y)$ is shown in Equation (9):

$$P\,(I_t) = p\,(x_{k,t}|x_{k,1}, ..., x_{k,t-1}) = \sum_{k=1}^{K} \omega_{k,t} f \left(I_t|\mu_{k,t}, \sum_{k,t} \right) \tag{9}$$

We try to sort K Gaussian mixture models from large to small by the weight $\omega_{k,t}$ and the priority $P_i = \frac{\omega_{i,t}}{|\Sigma_{i,t}|^{\frac{1}{2}}}$. Usually, the background model is built by the first B Gaussian distributions, as shown in the Equation (10):

$$B = \text{argmin} \left| \frac{\sum\limits_{k=1}^{b} \omega_k}{\sum\limits_{k=1}^{K} \omega_k} > T \right| \tag{10}$$

We match the current pixel I_t and the K sequential Gaussian by order. If I_t and a Gaussian distribution satisfy Equation (11), I_t matches this Gaussian distribution.

$$|I_t - \mu_{i,t-1}| \leq D_1 \delta_{i,t-1} \tag{11}$$

Among them, $\mu_{i,t-1}$ expresses the average of the i-th Gaussian function at $t-1$. The variable $\delta_{i,t-1}$ expresses standard deviation of the i-th Gaussian function at $t-1$. For the custom constant D_1, the value is usually 2.5. After background model matching, we establish the Gaussian mixture background model.

4.3. Updating the Gaussian Mixture Background Model

In order to improve the typical Gaussian background model algorithm, this paper adjusts several important parameters of the Gaussian background model. One of the objectives is to find the best match for each parameter. The other objective is to establish an efficient Gaussian mixture background model.

(1) Adjust the number of components of the Gaussian model: According to the experience, the number of Gaussian model components is between two and five. This value varies according to the complexity of the environment. If the value is too small, the Gaussian mixture distribution will be reduced to the single Gaussian distribution. Then, the Gaussian mixture background model is useless and cannot adapt to the complex background. If the value is too large, the computational load will be greatly increased and it will reduce the computational speed.

(2) Adjust the mean square deviation threshold of the Gaussian background model. The mean square deviation threshold is a criterion to determine the matching relationship between Gaussian components and the current pixel. The mean square deviation threshold should be selected according to the different environments. If the value is too small, it will make the system too sensitive and make a misjudgment. If the value is too large, the decision conditions will fail. The system will always believe that the original Gaussian matching matrix is the best matched matrix. This background model cannot detect the moving target effectively.

(3) Adjust the update rate of the Gaussian model. The update rate α is between 0 and 1. It shows that the Gaussian model is adapted to the current frame rate. If the value of α is too small, the initial modeling speed will be slow. If the value of α is too large, the initial modeling speed will be fast but it may reduce the inhibitory effect of the model on noise.

The defects of the typical updating algorithm are as follows: the update rate α will affect the initial modeling speed of the whole Gaussian mixture background model. If α's value is too small, the Gaussian model cannot construct the background that the current frame needs and takes a longer time for background modeling. On the contrary, if α's value is too large it may reduce the inhibitory effect of the Gaussian model on noise.

The improved Gaussian mixture model uses a self-adaption update rate α in the first N frames. The update rate α is larger and can change in real time. Therefore, initializing some frames will increase the influence on the whole model. The update rate α is determined by Equation (12):

$$\alpha = \begin{cases} 1/t \ \text{ if} : t < N \\ \alpha' \quad else \end{cases} \tag{12}$$

In the equation, N is a constant expressing the first N frames of the video, while t expresses the number of current flow frames in the video, and α' expresses the update rate of the typical Gaussian mixture model algorithm. N should be selected appropriately. If the value of N is too small, the statistics of gray values are too little and the initially established background model cannot reach the stable Gaussian distribution. If the value of N is too large, the update rate α' should be greater than or equal to the update rate α' after N frames. Then, $1/N$ is greater than or equal to the update rate α', and we can gather that N is less than or equal to the value of $1/\alpha'$.

4.4. Detection of a Moving Human Target

After updating the Gaussian mixture model, we arrange every Gaussian distribution by the size of priority $\frac{\omega_{i,t}}{\delta_{i,t}}$. T is the threshold value of the background weight. If the sum of the first B Gaussian distributions' weight is larger than T, the first B Gaussian distributions will be distributed as the background, and the others will be distributed as the foreground.

Human movement target detection subtracts the background model from the current video image. In this process, the pixels of a region that are different from the background model will be detected. This is the moving target region.

We match the current pixel I_t with each Gaussian distribution according to the order of priority. If both I_t and the background distribution satisfy Equation (13), we determine the point as a foreground pixel, or conversely a background pixel.

$$|I_t - \mu_{i,t}| > D_2\delta_{i,t}, i = 1, 2, ..., B \tag{13}$$

Among them, D_2 is a custom constant, and B is the number of the background distribution.

4.5. Post-Treatment

After detecting the movement target region, we process the detection region using morphological methods. Usually, the movement target region includes the foreground and areas of environmental noise. So, by the processes of erosion and dilation described above, we remove isolated foreground points and areas of noise. At the same time, we fill the foreground of the object image. Thus, finally, we detect the movement of the human target.

In morphology, the purpose of the erosion is to eliminate the object's boundary points. The erosion operation causes the boundary of the target object to shrink into the interior. This will remove some points which are not structural elements and are useless. The effect of erosion algorithms in this paper is to eliminate isolated foreground points. The mathematical expression is shown in Equation (14):

$$S = X \otimes B = \left\{ x, y \middle| B_{xy} \subseteq X \right\} \tag{14}$$

The set S expresses the binary image after erosion. The structure element B is used for erosion and can be composed of any shape. The value of each element in the structure element is 0 or 1. There is a central point in the structure element B. The set X expresses the pixel collection of the original image after the binary processing. Equation (22) describes the erosion of X by structure element B to get S. The set S is formed by all elements of B which are completely contained in X.

The judgement on whether a point is removed or retained proceeds as follows: The central point of the structure element B translates to the point (x, y) on the image X. If all pixels in B are identical to the corresponding pixels at point (x, y) as the center, the pixel (x, y) is retained. Remove the image pixels that do not meet the conditions. In this way, the boundary of the object will shrink into the interior.

The effect of the dilation operation and the corrosion operation are opposing: The dilation operation takes place in order to expand the object boundary points. After dilation, two disconnected objects may become connected. In this paper, the function of dilation is filling the foreground of the object image. The mathematical expression of this operation is shown in Equation (15):

$$S = X \otimes B = \left\{ x, y \middle| B_{xy} \cap X \neq \varnothing \right\} \tag{15}$$

Among them, the set S expresses the binary image after dilation. The structure element B is used for dilation. The value of each element in the structure element is 0 or 1. It can compose any shape. There is a central point in the structure element B. The set X expresses the pixels of the original image which are after the binary processing.

5. Human Object Contour Extraction Method Based on the Sobel Edge Detection Operator

At present, the often used contour extraction methods are: the method based on edge detection operator; the method of the threshold segmentation; the method of active contour modeling; and the method of dynamic programming graph search. In this paper, we use the edge detection method based on the Sobel edge detection operator to detect the moving object contour.

The edge of an image has two attributes, direction and magnitude. The changes of pixels along the edge direction are flat, and the changes of pixels perpendicular to the edge are more intense. The Sobel edge detection operator algorithm is relatively simple and calculates fast. However, the shortcomings of the Sobel edge detection operator algorithm are obvious. It can be clearly seen it is only a two direction template: It can only detect the horizontal and vertical edges. Therefore, the detection effect of an image with a complex texture is not good. In order to improve the accuracy of detection results, this paper increases the template to six directions based on the typical Sobel algorithm. The six direction templates are 45°, 135°, 180°, 225°, 270°, and 315° and are shown in Figure 1. The edge directions that can be detected by the eight direction templates are shown in Figure 2.

$$\begin{bmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

45° 135° 180°

$$\begin{bmatrix} 2 & 2 & 0 \\ 2 & 0 & -1 \\ 0 & -1 & -2 \end{bmatrix} \quad \begin{bmatrix} 0 & -1 & -2 \\ 2 & 0 & -1 \\ 2 & 2 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

225° 315° 270°

Figure 1. Adding the six directions of the template.

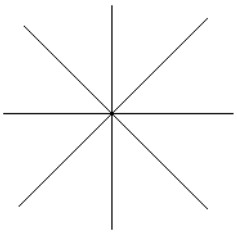


Figure 2. Detecting eight edge directions.

We use these eight direction templates to move on the image $f\ (x, y)$, respectively. We perform the neighborhood convolution for each pixel. The maximum value of the eight calculation results is the gradient value of this pixel point. The direction of the template corresponding to the maximum value is the edge direction of the point.

In practical applications, because of noise that produces edge discontinuity, the edge we extract is usually not closed. In this paper, we use a morphological dilation algorithm to make the human body contour edge connection.

6. Results

In order to verify the feasibility and effectiveness of our method in this paper, we have designed and implemented our method. We also have used our method in practical projects. Our system hardware consists of an infrared camera and computer peripherals. The computer hardware environment is comprised of an Intel Core2 duo 2.8 GhZ dual-core CPU, 4 GB of RAM, and an NVIDIA GeForce G210 video card. The system software environment for the computer system is Windows 7; the system running environment is Microsoft Visual Studio 2012 (*Visual Studio*, 2012; Microsoft, Redmond, WA, USA) and OpenCV (*OpenCV 2.4.8 for Windows*, 2013; Willow Garage, CA, USA).

6.1. The Experiment of Human Posture Recognition Method Based on HMM

Before we designed our system, we establish a new human posture database to determine and train the optimal system parameters.

(1) In the experiment, we have chosen four key postures for each behavior artificially. We use B, M, L, and E to name these four key positions. Therefore, the new human behavior database table size is 40.
(2) The transition probabilities for each state are as follows: $a_{BB} = a_{BM} = 0.5$, $a_{MM} = a_{ML} = 0.5$, $a_{LL} = a_{LE} = 0.5$, $a_{EE} = 1$.
(3) The probabilities of the initial state settings are: $\pi_B = 1$, $\pi_M = 0$, $\pi_L = 0$, $\pi_E = 0$.
(4) In general, a short position sequence is not sufficient to characterize a behavior, and if the sequence is too long, it will reduce the generalization performance of the behavior. Thus, in this experiment, we set up the posture sequence of the length T as 20 for all training and test sets. The specific parameters are initialized as shown in Table 1:

Table 1. The initialization of the Hidden Markov Model (HMM) parameters.

Parameter Name	**Definition or Initial Value**
Recognition target	An unknown behavior sequence
Number of HMM chain state	4
The initial state probability π	[1,0,0,0]
The initial state transition probability A	[0.5 0.5 0 0; 0 0.5 0.5 0; 0 0 0.5 0.5; 0 0 0 1]
The initial output probability B	$[1/40, 1/40, \ldots, 1/40]$
Position sequence length T	20

In our human posture database, we define about 10 typical postures, including "walk", "hands up", "skip", "raise the left hand", "raise the right hand", "run", "dash", "hands down", "stretch hands forward", and "turn around". So, we designed our method accordingly and tested the average recognition rate in interactive real time. After much training and testing, the experimental results show that the method proposed in this paper can produce a 92.75% average recognition rate.

As Table 2 explains:

(1) The y-axis represents the test set of each behavior, and each test set has 40 test datasets. Each dataset has four frames. The x-axis represents each corresponding posture of the HMM.

(2) The data on the table are the average recognition results repeated 10 times. A1–A10 represent the postures: “walk”, “hands up”, “skip”, “raise the left hand”, “raise the right hand”, “run”, “dash”, “hands down”, “stretch hands forward” and “turn around”.

(3) The number of data points on the diagonal represents the correct identifications, and the number of data points that are not on the diagonal is the number that were incorrectly classified.

(4) A clear confusion occurs between the behaviors “skip (A3)”, “run (A6)” and “dash (A7)”. We think this is normal, because the “skip” silhouette graph is very close to that of “run” and “dash”. Even people find these two behaviors hard to distinguish.

Table 2. The confusion matrix for our human posture recognition based on HMM in our database.

	A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**A10**
A1	37	0	0	1	0	0	1	0	1	0
A2	0	39	0	0	0	0	0	0	0	1
A3	0	0	35	0	1	2	2	0	0	0
A4	0	0	0	40	0	0	0	0	0	0
A5	0	0	0	0	40	0	0	0	0	0
A6	1	2	0	1	1	33	0	1	1	0
A7	0	0	1	0	2	0	35	0	1	1
A8	0	1	0	1	0	0	0	38	0	0
A9	0	1	0	0	2	0	0	0	36	1
A10	1	1	0	0	0	0	0	0	0	38

In this paper, we designed a variety of different actions. Each action is assigned to four key states in the HMM. Accordingly, $M = 4$ N, i.e., the number of HMM chain states is four. These four states are the “Beginning”, “Middle”, “Late”, and “End” states. In the experiment, we express these states with “B”, “M”, “L”, and “E”. Then, we break down the posture into four specific gestures. For example, “Raise the left hand” (A4 in Table 2 and Figure 3a) can be decomposed into four steps, including “standing”, “slightly raised left hand”, “left hand nearly horizontal”, and “left hand horizontal”. The corresponding initial state probability π is [1,0,0,0]. The corresponding initial state transition probability A is [0.5 0.5 0 0; 0 0.5 0.5 0; 0 0 0.5 0.5; 0 0 0 1]. The initial output probability B is $[1/40, 1/40, \ldots, 1/40]$. The corresponding position sequence length T is 20.

So, in our system we gained a different series of observation sequences about the behavior of the human movement. Different observation sequences showed different postures of the human body. We defined each posture of the observation sequence as a state. When the people in the scene made a movement, we calculated the probability and matched the state corresponding to the maximum probability with the observation sequence. Then, we classified the human motion, thus achieving our purpose of human posture recognition. Figure 3a shows a sequence of people extending their left arm. Figure 3b shows a sequence of people extending the right arm. Figure 3c shows a sequence for the pause indication. Figure 3d shows the motion that indicates the restart of a motion sequence.

In this experiment, the specific application of the system in regards to recognition and accuracy can achieve a higher level. In the test, more than 1000 times, the recognition of the correct rate of an easy posture like “Raise the left hand” (A4 in Table 2) or “Raise the right hand” (A5 in Table 2) can reach 100%. For other common actions such as “Hands up” (A2 in Table 2) or “Hands down” (A10 in Table 2), it can achieve more than 95% recognition of the average correct rate. The correct rate of recognition for all gestures can reach 92.75%. This is satisfactory for normal applications. As shown in the example in Figure 4a, when the participant extends the left arm, the tank on the screen turns left. As shown in Figure 4b, when the participant extends the right arm, the tank on the screen turns right. When the tank received the suspension of movement instruction issued by the human body, the tank stopped moving, as shown in Figure 4c. When the tank stopped and received restart instructions, the tank started moving again, as shown in Figure 4d.

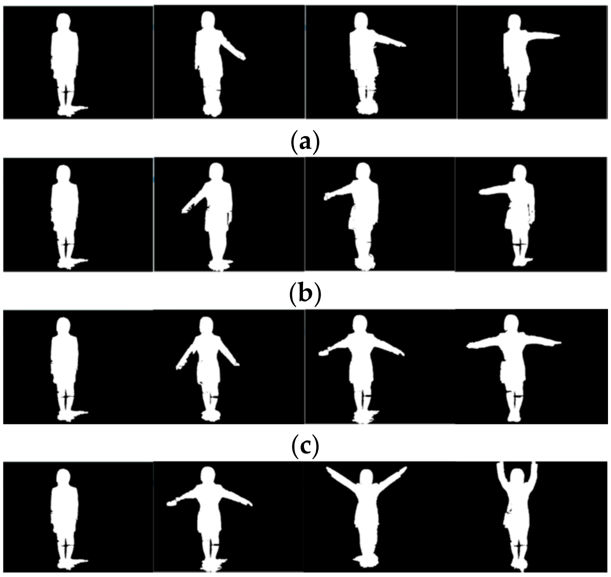

(**a**)
(**b**)
(**c**)
(**d**)

Figure 3. The sequences of extending or retracting the arm. Observation sequence of (**a**) extended left arm; (**b**) extended right arm; (**c**) pause indication; (**d**) restarting motion.

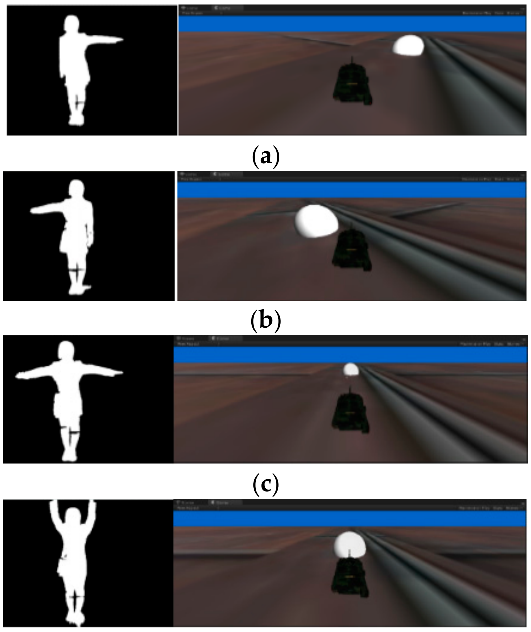

(**a**)
(**b**)
(**c**)
(**d**)

Figure 4. The tank on the screen responds to the instructions made by human body movement. (**a**) Detection of the human body extending the left arm: the virtual tank turns left; (**b**) Detection of the human body extending the right arm: the virtual tank turns right; (**c**) Detection of the human body pausing to indicate that the running tank has stopped; (**d**) Detection of the human body giving start instructions and tank beginning to run again.

6.2. The Experiment of the Human Movement Target Detection

To verify the research based on the improved Gaussian mixture background model subtraction, we designed and implemented our method and tested it in applications, in order to detect a moving human target in a scene.

When people enter a scene, the scene image will differ from the previous background image. After the background subtraction, we can detect a region that distinctly differs from it. This is the human moving target area. After detecting the target area of the moving human body, we can obtain the gray image. By setting a certain threshold value, we detect the target area of the movement of the human body, and binarize the area values. In this manner, we get a more obvious image of the movement of the human body. Figure 5a shows the infrared camera image of a person entering the scene. The human motion target area can be identified after the background subtraction is performed. Figure 5b is the video image after the according Gaussian mixture background model subtraction, showing the detected moving human target area. Figure 5c shows the binarized image of the movement target area.

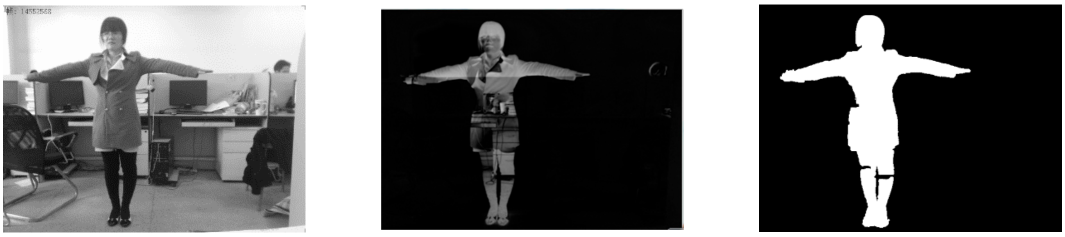

(**a**) (**b**) (**c**)

Figure 5. Image of a person entering the scene at different stages of the process. (**a**) Image of a person entering the scene; (**b**) Image after background model subtraction; (**c**) Binarized image, detecting the movement target area.

Due to the limitations of the detection method, noise, and threshold selection, the test results may not be ideal. Therefore, we need to continue the image detection post-processing. This concerns morphological processing, i.e., erosion of image points and dilation of prospective target image points, thereby realizing the ultimate goal of human movement detection. Figure 6a shows the image of the moving target area after erosion. We can see that the human object boundary shrinks inwards, against the isolated boundary or the erosion background. Figure 6b shows the image of the moving target area after dilation. Figure 6c shows the typical Gaussian mixture background model testing, whose noise resisting abilities are poor by comparison. As shown in Figures 5 and 6, we can clearly find that in Figure 6c, displaying the traditional method, noise is greater, and human body outline is more incomplete, than in Figure 5c, which displays our improved method. Thus, the traditional method cannot live up to application standards, whereas our improved method shows reduced noise after dilation and erosion processing, as shown in Figure 5c. With this method, the human body outline is relatively complete. Thus, by comparison, this detection effect is better.

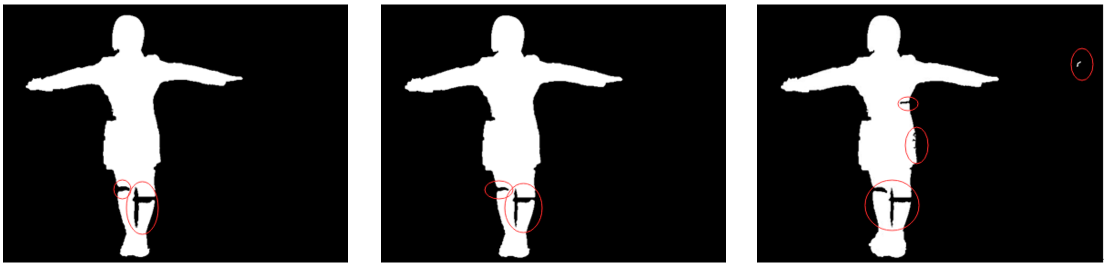
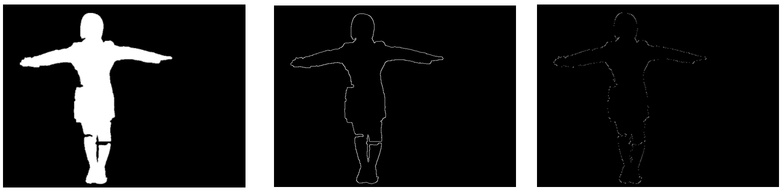

(**a**) (**b**) (**c**)

Figure 6. The various stages of image processing by the traditional Gaussian mixture background model. (**a**) Image of moving target area after erosion; (**b**) Image of moving target area after dilation; (**c**) Typical images after traditional Gaussian mixture background model testing.

6.3. The Experiment of the Human Object Contour Extraction

In order to verify the algorithm of edge detection, based on improving the Sobel edge detection operator to detect the feasibility and effectiveness of the moving human target contour extraction, we designed and then applied this method.

When using an improved Sobel edge detection operator, first we need to dilate the object, so that the contour edges are not connected, and the moving human target contour is more complete. Figure 7a shows the image of the moving human body after the according dilation operation: The edge of the contour expands, so that the contour edge is connected. The template is based on 8 directions of human movement target contour detection, improving the algorithm by adding 6 directions the typical Sobel edge detection operator algorithm that is the basis for the template. The improved algorithm achieves 8 directions. After the training, we can obtain a contour image as shown in Figure 7b. By comparing it with images generated by the typical Sobel edge detection operator, as shown in Figure 7c, we find that the 8 directions of the improved template are more complete and that the according contour obtained is more convenient to use. So, when applying the improved method, we can obtain a clearer, more complete contour that is easier to use



(**a**) (**b**) (**c**)

Figure 7. Comparison of typical Sobel edge detection operator and improved edge detection contour. (**a**) The moving human body after the dilation operation; (**b**) Contour of improved edge detection; (**c**) Contour of the typical Sobel edge detection operator.

6.4. Cloud-Based Monitoring System for the Elderly Using Our Method

We also developed a cloud-based monitoring system for the elderly using our method. In our system, we have three modules: posture detection, posture recognition, and monitoring visualization, shown in Figure 8. In our posture detection module, the posture images are captured by an infrared camera, which is installed on a distributed computer in an old person's room. After the postures in the images are detected by the client's computer, the datasets of the computation results are uploaded to the cloud storage, including the client's IP address and the posture recognition result. Using the cloud storage, the system provides a data visualization interface to display a global distribution status of the clients intuitively. In this way, administrators can monitor the old person in real time if an emergency situation arises.

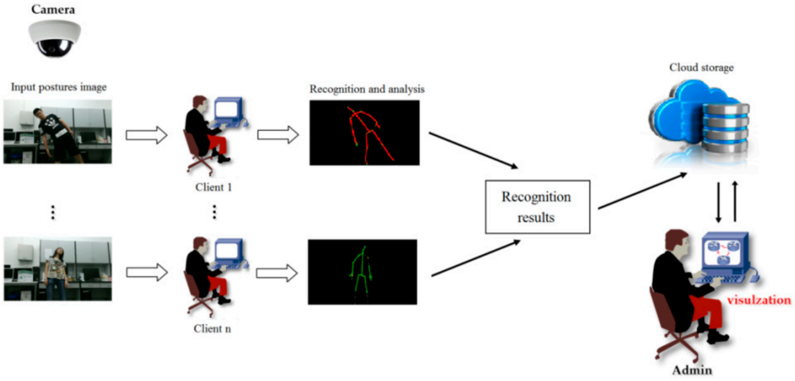


Figure 8. Our cloud-based monitoring system for the elderly.

In the data visualization module shown in Figure 8, we implement the data visualization technology on a web browser, which depicts a geographic location to provide a global distribution of the posture detection and recognition results. In the cloud storage, the client's IP address is translated into geographic information. Based on the converted geographic datasets, the data computed by the distributed client are localized onto the global map, which can be searched by Google Maps or OpenStreetMap (OSM, 2006; Steve Coast, CA, USA) in this system [26,27].

So, our human posture recognition system for smart home monitoring can monitor the behavior of an elderly person in real time. Using this system, we can automatically recognize when a person experiences sudden illness and can contact a hospital on their behalf. Also, a doctor can examine the patient's behavior, which assists in the preparation of the needed medical treatment.

We also have designed and implemented an infrared stage performance system using our method. In our infrared stage performance system, we set an infrared light source up so as to view the entire stage. We used the infrared camera to detect the human moving target, and detected the movement of the human body contour using the Sobel edge detection operator. After that, we extracted the human object contour. Then, we were able to use a person's contours and postures in relation to touching different-colored rectangles that represented different musical notes respectively. In this way, we were able to play musical games. Also, the different-colored rectangles could change to represent various English words: The participant could then touch the different words to hear their pronunciation. As our system can detect many persons and process the contour of each person, our interactive musical games also worked as multi-player games.

7. Conclusions and Future Work

With the rapid development of smart homes, we need a human posture recognition system for monitoring the behavior of elderly people. In this paper, we provide an infrared human posture recognition method for sustainable smart home monitoring based on a HMM. We built the human posture models based on HMMs and defined the model parameters. This model can be used to effectively classify the human postures.

Compared with the traditional HMM, this paper puts forward a method which solves the problem of human posture recognition. This paper tries to establish a model for treating data according to the characteristics of human activities. According to the characteristics of human postures, a complex problem can be deconstructed, thereby reducing the computational complexity. In practical applications, it can improve system performance.

Through experimentation in an actual environment, it was shown that in practical applications the model can identify the action of different body postures by observing human posture sequences, as well as matching recognition and classification processes. In addition, for human movement target detection, this paper puts forward a method of human movement target detection based on the Gaussian mixture model. For human object contour extraction, this paper puts forward a method of human object contour extraction based on the Sobel edge detection operator. Also, we propose an improved method after the analysis of the shortcomings of the original algorithm.

For the human movement target detection method based on the Gaussian mixture model, we carried out the following work: The first step was to capture the video image median filtering and to reduce the noise effect on the test results. The second step was to establish the Gaussian mixture background model. The third step was to combine the video image and the background model subtraction to detect the moving target area. Finally, we detected the moving target areas through the corrosion expansion processes. This increased the accuracy of the test results greatly. For the traditional Sobel edge detection operator, the moving object contour extraction method only detected the horizontal and vertical directions. An improved method was proposed to build on the six directions of the template.

Thus, we presented an experiment for human posture recognition, and we also examined a cloud-based monitoring system for elderly people using our method. We have tested our method in experiments, and the results show that our method is feasible and effective.

In future work, we will consider using multiple cameras to capture more complex human actions. It can improve the effectiveness of this method in practical applications. In addition, the actions of two or more people should be captured and a model established. This will be another focus of future research. Also, we will take into perspective applications of smart home monitoring with regard to other abnormal behaviors, e.g., also in other age groups, to avoid unnecessary health risks. We can use human posture recognition to operate variously in smart home without traditional remote controls. Detecting abnormal situations such as an old or disabled person falling down at home can be used to warn the family and/or to inform the hospital in a timely manner.

Acknowledgments: This research was supported by the Funding Project of National Natural Science Foundation of China (No. 61503005), the Funding Project of National Natural Science Foundation of Beijing (No. 4162022), the Funding Project of National Science and Technology Support Plan (2012BAF84F02), and the Funding Project of Great Wall Scholar Reserve Training Program in North China University of Technology (No. NCUT2015006). We would like to thank those who care for this paper and our projects. Also, we would like to thank everyone who spent time reading earlier versions of this paper, including the anonymous reviewers.

Author Contributions: All the authors have co-operated for the preparation of this paper. Xingquan Cai, Yufeng Gao and Wei Song conceived and designed the experiments. Xingquan Cai, Yufeng Gao and Mengxuan Li performed the experiments and analyzed the data. Yufeng Gao and Wei Song drafted the main part of the paper. Wei Song performed final reviews, including final manuscript rectifications.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Andre, E. Exploiting unconscious user signals in multimodal human-computer interaction. *ACM Trans. Multimed. Comput. Commun. Appl.* **2013**, *9*, 96–110. [CrossRef]
2. Alepis, E.; Virvou, M. Multimodal object oriented user interfaces in mobile affective interaction. *Multimed. Tools Appl.* **2012**, *59*, 41–63. [CrossRef]
3. Hasan, H.; Sameem, A.-K. Human-computer interaction using vision-based hand gesture recognition systems: A survey. *Neural Comput. Appl.* **2014**, *25*, 251–261. [CrossRef]
4. Riener, R.; Novak, D. Movement onset detection and target estimation for robot-aided arm training. *aT-Automatisierungstechink* **2015**, *63*, 286–298. [CrossRef]
5. Zelinsky, G.-J.; Peng, Y.; Berg, A.-C. Modeling guidance and recognition in categorical search: Bridging human and computer object detection. *J. Vis.* **2013**, *13*, 1–20. [CrossRef] [PubMed]
6. Kant, K.; Midkiff, S.-F. Pervasive computing and communications for sustainability. *Pervasive Mob. Comput.* **2012**, *9*, 118–119. [CrossRef]
7. Mahapatra, A.; Mishra, T.-K.; Sa, P.-K. Human recognition system for outdoor videos using Hidden Markov model. *AEC Int. J. Electron. Commun.* **2014**, *68*, 227–236. [CrossRef]
8. Satpathy, A.; Jiang, X.-D.; Eng, H.-L. LBP-based edge-texture features for object recognition. *IEEE Trans. Image Process.* **2014**, *24*, 1953–1964. [CrossRef] [PubMed]
9. Raman, N.; Maybank, S.J. Action classification using a discriminative multilevel HDP-HMM. *Neurocomputing* **2015**, *154*, 149–161. [CrossRef]
10. Chakraborty, B.; Bagdanov, A.-D.; Gonzalez, J. Human action recognition using an ensemble of body-part detectors. *Expert Syst.* **2013**, *30*, 101–114. [CrossRef]
11. Takano, W.; Obara, J.; Nakamura, Y. Action recognition from only somatosensory information using spectral learning in a Hidden Markov model. *Robot. Auton. Syst.* **2016**, *78*, 29–35. [CrossRef]
12. Yao, B.; Hagras, H.; Alhaddad, M.-J. A fuzzy logic-based system for the automation of human behavior recognition using machine vision in intelligent environments. *Soft Comput.* **2015**, *19*, 499–506. [CrossRef]
13. Nakamura, E.; Nakamura, T.; Saito, Y. Outer-product Hidden Markov model and polyphonic midi score following. *J. New Music Res.* **2014**, *43*, 183–201. [CrossRef]
14. Azzam, R.; Kemouche, M.-S.; Aouf, N. Efficient visual object detection with spatially global Gaussian mixture models and uncertainties. *J. Vis. Commun. Image Represent.* **2016**, *36*, 90–106. [CrossRef]

15. Singh, S.; Saini, A.-K.; Saini, R. A novel real-time resource efficient implementation of Sobel operator-based edge detection on FPGA. *Int. J. Electron.* **2014**, *101*, 1705–1715. [CrossRef]
16. Guha, T.; Ward, R.-K. Learning sparse representations for human action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *38*, 1576–1588. [CrossRef] [PubMed]
17. Ahmad, M.; Lee, S.-W. Human action recognition using shape and CLG-motion flow from multi-view image sequences. *Pattern Recognit.* **2008**, *41*, 2237–2252. [CrossRef]
18. Zhao, H.-Y.; Li, C.-Y. Human action recognition based on multi-features fusion. *J. Appl. Res. Comput.* **2012**, *29*, 3169–3172.
19. Du, B.; Geng, X.-L.; Chen, F.-Y.; Pan, J.; Ding, Q. Generation and Realization of Digital Chaotic Key Sequence Based on Double K-L Transform. *Chin. J. Electron.* **2013**, *22*, 131–134.
20. Wu, Q.-X.; Deng, F.-Q.; Kang, W.-X. Human action recognition in complex scenes based on fuzzy integral fusion. *J. South China Univ. Technol.* **2012**, *40*, 146–151.
21. Chen, W.-Q.; Xiao, G.-Q.; Lin, X.; Qiu, K.-J. On a human behaviors classification model based on attribute-bayesian network. *J. South China Univ. Technol.* **2014**, *39*, 7–11.
22. Liang, F.; Zhang, Z.-L.; Li, X.-Y.; Tong, Z. Action recognition of human's lower limbs in the process of human motion capture. *J. Comput. Aided Des. Comput. Graph.* **2015**, *27*, 2419–2426.
23. Dardas, N.-H.; Georganas, N.-D. Real-Time hand gesture detection and recognition using bag-of-features and support vector machine techniques. *IEEE Trans. Instrum. Meas.* **2011**, *60*, 2592–2607. [CrossRef]
24. Rabiner, L. A tutorial on hidden markov models and selected applications in speech recognition. *Proc. IEEE* **1989**, *77*, 257–286. [CrossRef]
25. Fink, G.-A. *Markov Models for Pattern Recognition*, 2nd ed.; Springer-Verlag: London, UK, 2008; pp. 71–106.
26. Hagenmeyer, V.; Cakmak, H.-K.; Dupmeier, C.; Faulwasser, T. Information and Communication Technology in Energy Lab 2.0: Smart Energies System Simulation and Control Center with an Open-Street-Map-Based Power Flow Simulation Example. *Energy Technol.* **2016**, *4*, 145–162. [CrossRef]
27. Kaklanis, N.; Votis, K.; Tzovaras, D. Open Touch/Sound Maps: A system to convey street data through haptic and auditory feedback. *Comput. Geosc.* **2013**, *57*, 59–67. [CrossRef]