International Journal of Smart Home
Vol.8, No.5 (2014), pp.207-218
http://dx.doi.org/10.14257/ijsh.2014.8.5.19

Human Action Recognition Algorithm: Risk Notification Service on the Android Environments

Nae-Joung Kwak[1] and Teuk-Seob Song[2]

[1]*Dept. Communication & Information, Chungbuk National University,*
[2]*Dept. Computer Engineering, Mokwon University,*
[1]*knj0125@hanmail.net*, [2]*teukseob@mokwon.ac.kr*

Abstract

This paper proposes a human action recognition algorithm which can be efficiently applied to a real-time intelligent surveillance system. The proposed method classifies human actions into walking, sitting, standing up and unusual action like faint and falling down. Also, in the case of detecting unusual actions, it offers an alarm to smartphones to monitor the object of interest. This method models the background, obtains the difference image between input image and the modeled background image, extracts the silhouette of human object from input image and recognizes human actions. In order to recognize human actions, the proposed method uses direction vector of movement and link codes of movement dependent histogram (LC-NMDH). Firstly, NMDH is computed by dividing the motion information histogram into ten parts and saving the median value of each part. LC-NMDH is defined as the values which records a chain code of NMDH of each part.

Human actions are classified into walking, sitting, standing up and unusual action. The proposed method was examined on seven people through sequences captured by a web camera. The proposed algorithm efficiently classified human actions, detected unusual actions and provided an alarm service. The test result showed more than 99 percent recognition rate for each action by the proposed method.

Keywords: *Risk Notification, Human Action Recognition, Histogram*

1. Introduction

There is a growing interest in establishing local and international safety nets and coming up with ways to deal with the dangers individuals face on a daily basis. To satisfy these needs, surveillance systems have been developed and installed everywhere. The previous surveillance systems using CCTV cameras require people to manage programs for maintenance and emergencies. Under this system, managers' judgment and decisions are absolutely crucial in an emergency. So their poor management can lead to huge economic losses and heavy damage.

An intelligent surveillance system can improve the problem of poor management by making the system analyze positions and patterns of the object and respond to each situation based on digitalized images sent from surveillance cameras [1]. As advances in scientific technology create ubiquitous environments, an intelligent surveillance system linked to home networks could build intelligent home networking systems forward. Such an intelligent home networking system is utilized to monitor infants, seniors living alone, and the physically challenged. Studies on human action recognition technologies are being conducted to detect

Corresponding Author: Teuk-Seob Song, Division of Convergence Computer and Media Mokwon University Daejeon Korea 302-318, E-mail: teukseob@mokwon.ac.kr

emergencies. Human action recognition technologies are applied to various fields such as video surveillance systems, human-computer interaction, video indexing, and sports video analysis. Human action recognition follows three steps: preprocessing, action modeling, and action recognition. In the preprocessing step, the object of interest is extracted and the background is subtracted [2]. Action modeling is a step to model human actions and get information needed for action recognition. For action modeling, there are two approaches: body structure approach and holistic approach. Body structure approach models human actions and then gains information needed to recognize actions. The holistic approach is based on the entire contour of the human body [3]. Body structure approach is divided into top-down [4] and bottom-up [5] approaches. Top-down processing first models body structure and matches input images with the modeling. Bottom-up processing first detects specific body parts from input images and connects them. This approach requires a lot of calculations because it has to detect each part of the body and estimate actions.

The holistic approach extracts bodies as objects and models actions by using form, contour, texture, silhouette, location, trajectory, and velocity of objects. Since this approach models bodies holistically and requires less calculations than body-part approaches, the holistic approach is widely used [6]. The features of extracted objects combined with occurring events are used as basic information to recognize human actions. Methods which extract objects of interest from video images and recognize their features and actions have steadily been studied and developed depending on image recognition [7-9]. However, the existing methods have to learn the characteristics extracted from human body data before they come to recognize and estimate human actions and postures. It means that they require a lot of training data and complex training algorithms.

The proposed method does not use training data for a faster recognition. It proposes an algorithm which can recognize usual actions like walking, sitting and standing up, and unusual actions such as fainting and falling down. Furthermore, in the case of unusual action, it sends an alarm message to smartphones to alert the person of interest.

Section 2 describes the means of object extraction, action classification, and alarm service. Section 3 evaluates the performance of the proposed method, and section 4 provides conclusions. Figure 1 is the system of the proposed method.

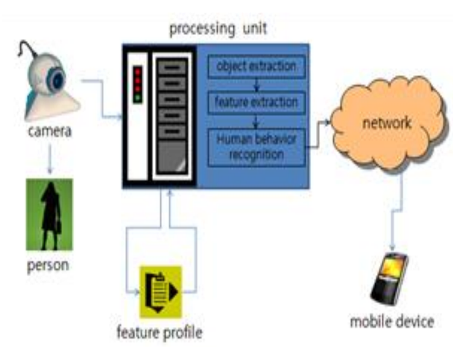


Figure 1. System of the Proposed Method

2. The Proposed Method

This paper proposes a method to recognize usual actions like walking, sitting, and standing up and unusual actions like faint and falling down, as well as service an alarm for the unusual actions. First, the proposed method models the background and separates objects from the background by using the difference between the modeled background and the input image captured by a single camera. The proposed method extracts objects using the background information. Then, it classifies human actions into four categories : walking, sitting, standing up, and unusual action by using a normalized motion dependent histogram (NMDH) and face movement trajectories.

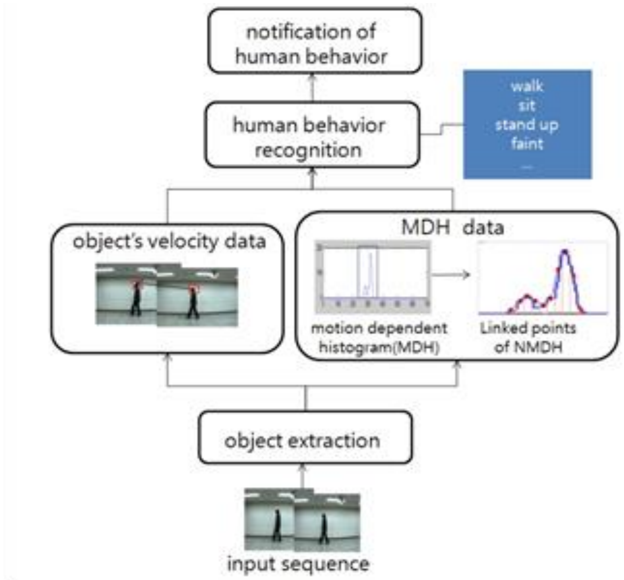


Figure 2. Flowchart of the Proposed Method

2.1. Background Modeling and Object Extraction

The proposed method extracts objects by using the method proposed in [10, 11]. The proposed method first obtains the difference image between the modeled background images and each input one. The difference image is an image where the background and the object are separated. If an input image has a pixel value between high and low thresholds on three planes, it is the background. Otherwise, it is an object. The following is the equation to separate objects from the background.

$$B(x,y)=\begin{cases}0 & , \quad Th_l \le I(x,y) \le Th_h \\ 255 & , \quad otherwise\end{cases} \tag{1}$$

Where x, y indicates the location of the input image. In the binary image, an object is white (255) and the background is black (0). By combining the results of three planes, the resulting image is produced.

The binary image has various segments which are not included in the object. Thus, we reduce segments and noises by using morphological filters and acquire the background-object separation image $B(x, y)$.

2.2. Parameter for Classifying Object's Actions

Human actions are recognized by the change in a human makes. For example, the upper portion of human body moves down for sitting, up for standing up, and sideways for walking.

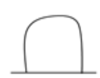
Therefore, in order to classify walking, sitting, standing up, and unusual actions such as falling down and fainting, the proposed method uses two parameters of NMDH and the direction and the velocity of human movement. The first parameter, NMDH is defined by normalizing the histogram of an object's movement for four actions. The other parameter, the direction and the velocity of human movement represents a change in one;s position. The proposed method detects the facial area of extracted objects and uses the direction and velocity of the movement of the center point in the area.

First, NMDH is acquired by using motion information histogram (MIH) in [10]. The values are computed by dividing the histogram of MIH (motion information histogram) into ten parts and saving the median value of each part. The proposed method calculates MIH by using motion information of objects and determines human actions by extracting MIH features of each action. Equation (2) is the equation of MIH.

$$MIH = (FB \bullet FC)\text{-}FB \qquad (2)$$

where FB indicates the background-object separation image of the previous frame. FC indicates the , background-object separation image of the current frame, and '• ', stands for "exclusive-or." MIH is a value defined only by objects' motion on the current frame when objects' movements are detected between previous and current frames. Table 1 shows MIH of four actions [10].

Table 1. MIH of Four Actions

Action	Walking	sitting	Standing up	Unusual
MIH				

To use NMDH as an action recognition parameter, the proposed method records a chain code of NMDH of each part according to Table 3 and defines LC-NMDH as the linked chain codes of NMDH.

As objects move, the location of an object's face changes. This paper detects objects' faces and tracks the changes of the facial area. The direction of the movement is used as an action classification parameter.

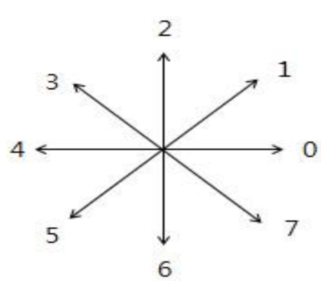


Figure 3. Chain Code

Table 2. LC-NMDH

Action	String of chain code of NMDH
Walk	1 – 1 – 1 – 1 – 7 – 7 – 7 – 7 - 7
Sit	7 – 7 – 7 – 7 – 0 – 0 – 1 – 1 - 1
Stand	1 – 1 – 1 – 0 – 0 – 0 – 0 – 0 –7
risk	1 – 1 – 1 – 0 – 0 – 0 – 7 – 7 - 7

We detect faces by looking for skin color information of the upper part of objects' bodies. Table 3 shows the changes of the center point in facial location with + and -. In Table 3, "*" indicates "irrelevant."

Table 3. Center point's Change of Facial Region for Action Classification

Action	Center point	
	X	Y
walking	+/-	*
sitting	*	+
standing	*	-
Unusual action	+/-	+

2.3. Object Action Classification and Unusual Action Recognition

The proposed method recognizes unusual actions first and then classifies the rest of the three actions. Figure 4 is a flowchart of classification of the four actions, and the following is the action recognition algorithm.

1. Detect the facial area in the input image. Extract the center point of the face and calculate the movement direction and velocity.

2. Calculate LC-NMDH in the object image.

3. Classify actions by calculating the value of process 2 and the distance between LC-NMHDs of Table 2.

4. If the velocity of face area of process 1 and the result of process 2 falls into the unusual action category, classify it as "Risk Situation."

5. Otherwise, classify the rest of the actions based on Table 3 and process 3

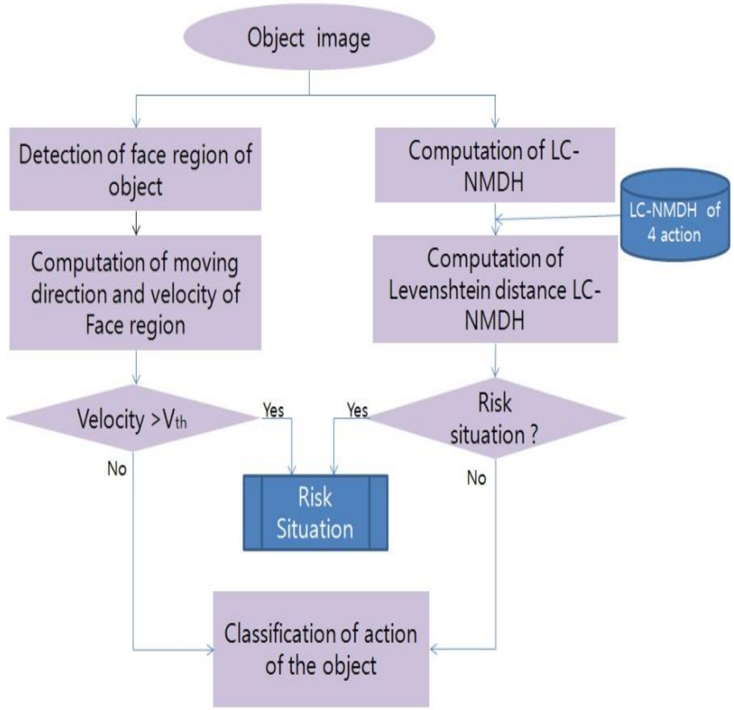


Figure 4. Flowchart of Classification of the 4 Actions of Object

The distances between LC-NMHDs of process 3 and LC-NMDHs of Table 2 are calculated by using Levenshtein distance [12]. The following is the Levenshtein algorithm to calculate the distance between strings.

```
Levenshtein distance computation
Output : distance d between x and y
        Corrected operating list edit list[]

Algorithm
1.   C: the length of x, r : the length of y
2.   Create the array D[0⋯r] [0⋯c] of (r+1)*(c+1)
3.   for(i=0 to c) D[0][i] = i;
4.       for(j=0 to r) D[j][0] = j;
5.           for(j=1 to r)
6.            for(i=1 to c) {
7.              if( xi = yi ) scost=0; else scost=1;
8.                   D[j][i] = mininum(D[j-1][i]+1, D[j][i-1]+1, D[j-1][i-1]+scost);
9.             act=the minimum operation in the line 8
10.            if( act='substitute' and scost=0)
                        act='no change';
```



```
            action[j][i]=act;
          }
    d=D[r][c] ;
     j=r, i=c;
    k=1;
  repeat{
          if (action[j][i]='insertion')
                {edit_list[k]='insertion';j--}
         else if (action[j][i]='deletion')
              {edit_list[k]='deletion';i--}
         else if (action[j][i]='substitution')
                {edit_list[k]='substitution';j--; i--}
        else if (action[j][i]='no change')
                {edit_list[k]='no change';j--; i--}
     k++;
   } until(j=0 and i=0);
    reverse the order in edit_list[]
```

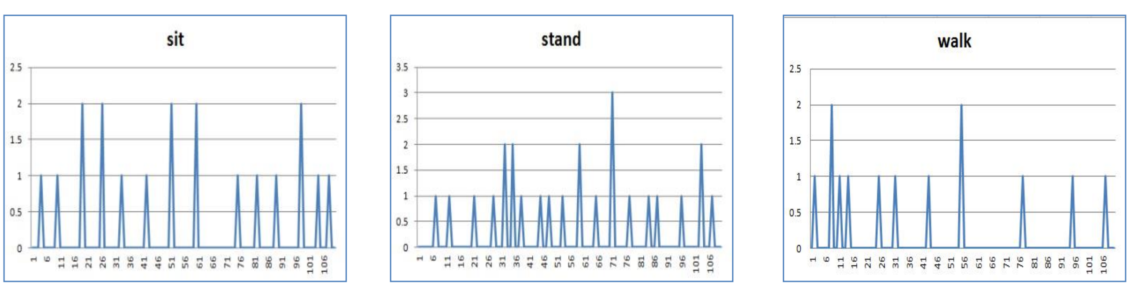
Figure 5. Levenshtein Algorithm

Table 4. Levenshtein Distance of Actions

actions	Walk-sit	Walk-stand	Walk-risk	Sit-stand	Sit-risk	Stand-risk
Levenshtein's distnace	9	5	5	9	9	2

Table 4 shows the combination of four actions and the Levenshtein distance between LC-NMDHs of actions.

Figure 6 shows the Levenshtein distance between LC-NMDHs of sitting, standing, and walking computed from input frames and reference ones form table 2. As for input 110 frames, there were 1 or 2 errors for each action. These errors were made by disagreement concerning ten parts of NMDH and the median values.



(a) Sit **(b) Stand** **(c) Walk**

Figure 6. Levenshtein Distance of each Frame for Each Action

According to Table 4, the Levenshtein distances between LC-NMDHs for three usual actions are higher than five. In the consideration of possible errors in calculating LC-NMDHs between objects, the Levenshtein distance between input frames and reference is below three. Therefore the proposed method classifies actions where the Levenshtein distance is lower than two.

2.4. Smartphone Alarm Service

When the proposed method detects an unusual action classified in Section 2.3, it notifies via smartphone to monitor the status of the object of interest. Figure 7 shows Server–Client mechanism for monitoring the object of interest. The server consists of connected modules, video input and storage modules, risk recognition modules, risk alarm modules, and video transfer modules. The client consists of the connection modules, risk signal monitoring and receiving modules, app-activating modules, and risk verification modules.

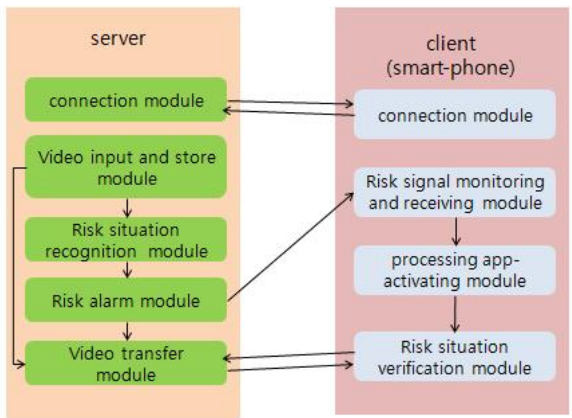
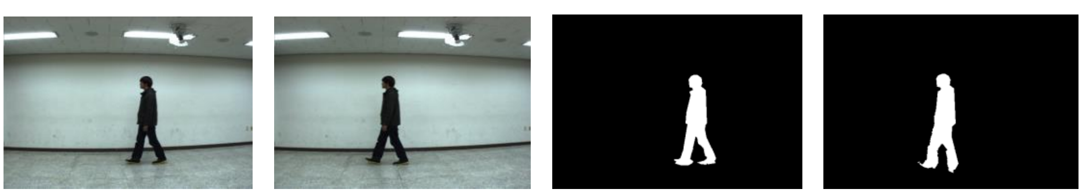


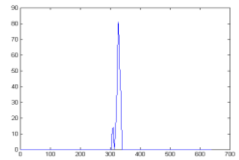
Figure 7. Server – Client Mechanism for Risk Monitoring of Human Object

3. Test and Results

This paper dealt with background and input images on a real-time basis with a camera to analyze the capacity of the proposed method. The resulting images show that each of the seven participants took four different actions such as sitting, walking, standing up, and falling/fainting against two different indoor backgrounds. Intel cpu (2.0GHz), 1G RAM, Visual Studio 2008, and OpenCV 2.1 were used. The resolution of the input image was 640×480 24 bit and the sequence was recorded at 15 frames/sec. The test of the risk notification service was conducted on Eclipse Android-based APIs level 8 in HVGA skin and SD card 32Mbyte.



(a) Walk Image (before) **(b) Walk Image (current)** **(c) Object Image of (a)** **(d) Object Image of (b)**



(e) MDH of (c) and (d)

(f) Motion of (c) and (d)

Figure 8. Result Images by the Proposed Method

Figure 8 shows the resulting images of the experiment. (a) and (b) are input frames of a walking action. (c) and (d) are the result to extract its silhouette and corner points after extracting a main object from images (a) and (b). (e) is NMDH of (a) and (b), and (f) is the trace result of a walking action

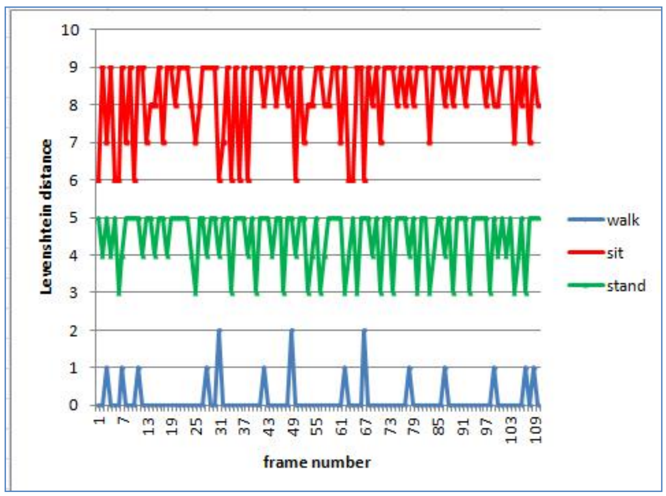


Figure 9. Levenshtein Distance For walk Action

Figure 9 shows the Levenshtein distances of LC-NMDH of walking, sitting, and standing up in 110 walking action frames. It demonstrates that Levenshtein distance precisely classifies actions.

Figure 10 (c) shows a screenshot of initial menu, , and Figure 10 (d) depicts a screenshot of the alarm service. Furthermore Figure 10 (e) portrays an image of the smartphone alarm service application.

(a) Walk Image (before)

(b) Risk Image (current)

(c) Initial Menu **(d) Alarm of Risk Situation** **(e) Video of Object**

Figure 10. The Entire Notification Process of the Proposed System

When the risk situation such as (b) is generated, the proposed method notifies the situation to the smartphone as in figure (d). If the user clicks the button for video display, the video of the object is displayed such as in image (e).

Table 5 shows the result of the proposed method examined on seven people with the number of frames recognized by the proposed algorithm. This is among 300 frames and the average recognition rate. Since unusual actions are suddenly taken, we do not limit the number of frames. The proposed method detects walking, sitting, and standing 99 percent of the time and unusual actions 100 percent of the time. This is because changes in lightning and wrong extractions due to noises lead to wrong detection of NMDH.

Table 5. The Result of Human Action Recognition

Action	Walking	Sitting	Standing	Unusual Action
1	300	298	300	-
2	298	300	298	-
3	298	296	300	-
4	300	300	300	-
5	299	296	298	-
6	296	300	300	-
7	300	295	299	-
Recognition Ratio	99%	99%	99%	100%

4. Conclusion

This paper proposes an algorithm that extracts object movement by using the difference image between background and input images captured by a single camera. The algorithm classifies objects' actions, provides an alarm service in risk situations, and monitors them at clients' requests. In order to efficiently extract objects, various backgrounds are modeled. We classify human actions into walking, sitting, standing up and unusual action like faint and falling down. As an action recognition parameter, the NMDH traces objects' facial areas.

Then, by using Levenshtein distance, the proposed method classifies actions based on velocity and direction of movement regarding facial areas. By combining two parameters, actions are finally classified.

Furthermore, the proposed method of the sequence captured by a web camera is examined. The proposed method successfully offers an alarm service of risk situations and shows more than a 99 percent detection rate for four actions. The proposed method does not require complex training data or algorithms and can have various applications such as surveillance cameras or u-Health in the ubiquitous environment.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2014-017928).

References

[1] R. T. Collins, A. J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt and L. Wixson, “A System for Video Surveillance and Monitoring”, Technical Report CMU-RI-TR-00-12, Carnegie Melon University, **(2000)**.
[2] J. Park, A. Tabb and A. C. Kak, “Hierarchical Data Structure for Real-Time Background Subtraction,” IEEE International Conf. on Image Processing, **(2006)**.
[3] J.K. Aggarwal and Q. Cai, “Human Motion Analysis: A Review,” Computer Vision and Image Understanding, Vol. 73, **(1999)**, pp. 428-440.
[4] R. Urtsasun, D. J. Fleet and P. Fua, “Monocular 3-D Tracking of the Golf Swing,” IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, **(2005)**, pp. 932-938
[5] D. Ramanan and D. A. Forsyth, “Finding and Tracking People from the Bottom up”, Proceedings of Computer Vision and Pattern Recognition (CVPR), vol. 2, **(2003)** June, pp. 467-474.
[6] Y. Alper and S. Mubarak, “Actions Sketch: A Novel Action Representation”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, **(2005)**, pp. 984-989.
[7] J. Tao and Y. P. Tan, “A Probabilistic Reasoning Approach to Closed-Room People Monitoring”, IEEE ISCAS, **(2004)**, pp. II-185-188.
[8] A. Utsumi, H. Yang and J. Ohya, “Adaptive Human Motion Tracking using Non-Synchronous Multiple Viewpoint Observations”, Proc. IEEE 15th International Conf. on Pattern Recognition, vol. 4, **(2000)**, pp. 607-610.
[9] T. E. de Campos and D. W. Murray, “Regression- based Hand Pose Estimation from Multiple Cameras”, CVPR 2006, vol. 1, **(2006)**, pp. 782-789.
[10] N. J. Kwak and T. S. Song, “Human Action Classification and Unusual Action Recognition Algorithm for Intelligent Surveillance System”, IT Convergence and Security 2012 Lecture Notes in Electrical Engineering, vol. 215, **(2013)**, pp. 797-804.
[11] N. J. Kwak and T. S. Song, “Human Action Recognition for Risk Notification Service”, ICFICE 2014, vol. 6, no. 1, **(2014)**, pp. 105-108.
[12] S. Giovanni, V. Kripasundar and R. K. Srihari, “Generalizing Edit Distance to Incorporate Domain Information: Handwritten Text Recognition as a Case Study”, Pattern Recognition, vol. 29, **(1996)**, pp. 405-414.

Authors

Nae-Joung Kwak received the B.S. in February 1993 and M.S. in February 1995, Ph.D in February 2005 from the department of Computer and Communication engineering, Chungbuk National University. Her research interests include multimedia communication, multimedia signal processing, video surveillance system, and MPEG.

Teuk-Seob Song received his Ph.D. degree in computer science in 2006 and Ph.D. degree in Mathematics from Yonsei University in 2001, respectively. He is currently an associate professor in the Division of Convergence Computer and Media at Mokwon University in Korea. His research interests include 3D Virtual Environment, Semantic Navigation, XML Document, Annotation Technology and Structured Document Transcoding. He is a member of the Korea Information Science Society, the Korea Information Processing Society, and the Korea Multimedia Society.