Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2015, Article ID 676090, 9 pages
http://dx.doi.org/10.1155/2015/676090

Research Article

Human Activity Recognition as Time-Series Analysis

Hyesuk Kim and Incheol Kim

Department of Computer Science, Kyonggi University, San 94-6, Yiui-Dong, Youngtong-Gu, Suwon-Si 443-760, Republic of Korea

Correspondence should be addressed to Incheol Kim; kic@kgu.ac.kr

Received 6 June 2015; Revised 26 August 2015; Accepted 27 August 2015

Academic Editor: Meng Du

We propose a system that can recognize daily human activities with a Kinect-style depth camera. Our system utilizes a set of view-invariant features and the hidden state conditional random field (HCRF) model to recognize human activities from the 3D body pose stream provided by MS Kinect API or OpenNI. Many high-level daily activities can be regarded as having a hierarchical structure where multiple subactivities are performed sequentially or iteratively. In order to model effectively these high-level daily activities, we utilized a multiclass HCRF model, which is a kind of probabilistic graphical models. In addition, in order to get view-invariant, but more informative features, we extract joint angles from the subject's skeleton model and then perform the feature transformation to obtain three different types of features regarding motion, structure, and hand positions. Through various experiments using two different datasets, KAD-30 and CAD-60, the high performance of our system is verified.

1. Introduction

Vision-based activity recognition has found many applications such as human-computer interaction [1, 2], surveillance [3, 4], robot learning [5, 6], and user interface design [7, 8]. Recently many researchers tend to use depth cameras like Microsoft Kinect to detect human activities. Unlike conventional RGB cameras, Kinect-style depth cameras can provide us with the depth information in addition to colors of the target object. Depth information can be used to estimate the 3D body poses of a human and to recognize his/her real-time activities. In this paper, we propose a system that can effectively recognize daily human activities with a Kinect-style depth camera. Our system utilizes a set of view-invariant features and the hidden state conditional random fields (HCRF) [9, 10] model to recognize human activities from the dynamic body pose estimates provided by MS Kinect API or OpenNI. Many high-level daily activities can be regarded as having a hierarchical structure, where multiple subactivities are performed sequentially or iteratively. Our system utilizes a multiclass HCRF model to represent effectively hierarchical nature of such activities.

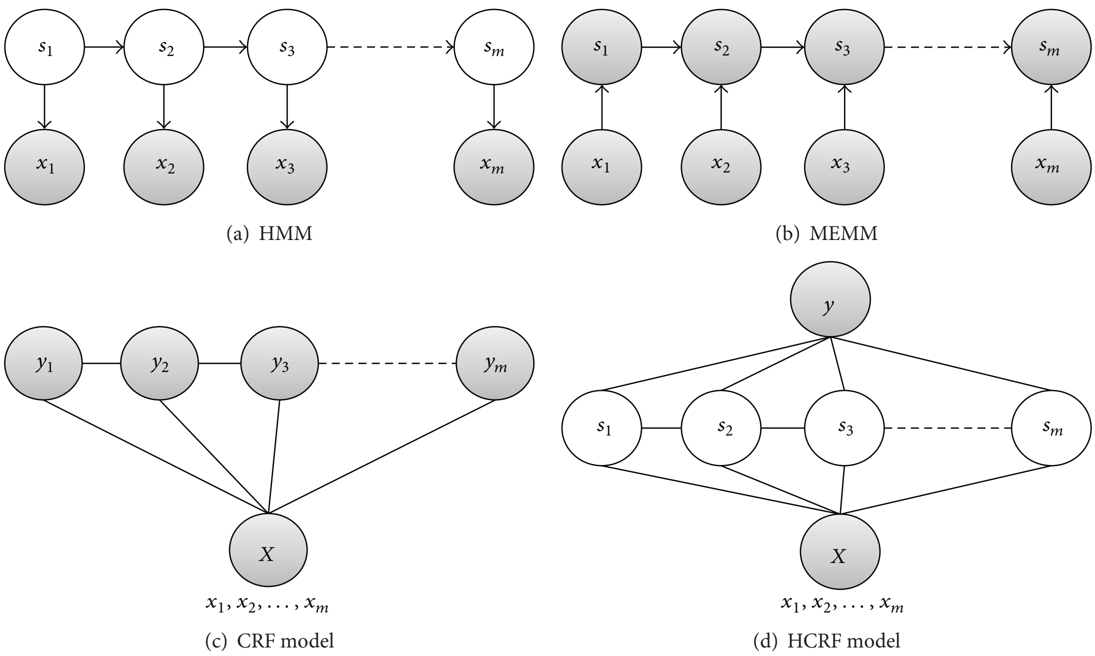
Many existing systems often make use of only 3D coordinates of individual body joints as a feature set for activity recognition. However, these joint coordinates can be affected easily by change of Kinect's viewpoint [11, 12]. In order to meet the view variance problem and get more informative features, our system extracts joint angles from the subject's skeleton model and then performs the feature transformation to get three different types of features regarding motion, structure, and hand positions.

The remainder of this paper is structured as follows. In Section 2, we briefly introduce the related works. Section 3 presents a comparison of various probabilistic graphical models including HMM, MEMM, CRF, and HCRF. Section 4 concentrates on the design of our activity system. Section 5 presents the conducted experiments using two different datasets and results obtained with our system. Finally, Section 6 summarizes our work and outlines the future work.

2. Related Works

The most important factors to affect the performance of vision-based activity recognition systems are both the set of features and the recognition model to capture the unique characteristics of individual activities. Previous works adopt different features and models from each other, resulting in distinct strength and weakness in performance.

In Xia et al.'s work [13], histograms were extracted from the joint coordinates as features using modified spherical



(a) HMM

(b) MEMM

(c) CRF model

(d) HCRF model

Figure 1: Probabilistic graphical models.

coordinate systems in order to overcome the view variance problem. However, for different activities that involve similar positions of the joints, the system could generate similar histograms, hence making it difficult to distinguish between the two activities. In this work, activities are modeled with Hidden Markov Model (HMM). The HMM is a widely used probabilistic graphical model to process a time-series data. However, this model has a limitation that current observations are only dependent on the current state, not on any previous states or observations. Moreover, it has another limitation on training efficiency since it requires supervised training to maximize the joint probability of observation and state sequences. On the other hand, in Sung et al.'s work [14], joint angles are used as features instead of the corresponding joint coordinates to overcome the view variance problem. The hierarchical Maximum Entropy Markov Models (MEMMs) are adopted to model the hierarchical nature of activities as well as enhance the training efficiency. However, MEMMs are well known to suffer from the label bias problem.

In Zhang and Tian's study [15], spatiotemporal features and Support Vector Machines (SVMs) were used to represent activities. However, the features do not consider the view variance problem and SVMs are limited in training human activity patterns over time in comparison with probabilistic graphical models. In Ong et al.'s work [16], features based on the human range of movement were extracted from joint poses and k-means clustering which is an unsupervised learning method is applied to recognize daily activities. However, the features of this work are sensitive to camera view variance and the range of motion of joints may vary from person to person. It recognizes activities through k-means clustering without training a model. However, k-means clustering has several limitations that the number of clusters should be predetermined and the resulting clusters may be varied depending on the given initial clusters as well.

3. Probabilistic Graphical Models

Probabilistic graphical models [17] can be considered as one of the best ways to represent hierarchical structures of high-level daily activities, where multiple subactivities are performed sequentially or iteratively. Among the widely used probabilistic graphical models for activity recognition are the Hidden Markov Model (HMM), the Maximum Entropy Markov Model (MEMM), and the Conditional Random Fields (CRF) as shown in Figures 1(a)–1(c), respectively.

The HMM in Figure 1(a) is a generative graphical model in which the target system to be modeled is assumed to be a Markov process. In the figure, the variables x_t, s_t, and y_t represent the observation, the hidden state, and the class label, respectively. This model assumes that the conditional probability distribution of the hidden variable s_t at time t depends only on the value of the hidden variable s_{t-1}. Similarly, it assumes that the value of the observation variable x_t only depends on the value of the hidden variable s_t. This means that the HMM presumes independence of the observations. Therefore, this model cannot represent long-range dependencies among observations. Additionally, it has another limitation on training efficiency since it requires supervised training to maximize the joint probability of observation and state sequences.

The MEMM in Figure 1(b) is a discriminative graphical model that combines the features of the HMM and the Maximum Entropy (MaxEnt) model. An advantage of MEMM over HMM is that it provides increased freedom in choosing features to represent observations. Another advantage of MEMM over HMM is that training can be considerably more efficient. In MEMM, estimating the parameters of the maximum-entropy distributions used for the transition probabilities can be done for each transition distribution in isolation. However, the MEMM has a drawback that it potentially suffers from the label bias problem, in which states with low-entropy transition distributions effectively ignore their observations.

The CRF model in Figure 1(c) is a discriminative undirected graphical model. In the figure, X represents the observation sequence and y_t represents the random variable which, conditioned on X, obeys the Markov property. The CRF model can contain any number of feature functions and the feature functions can inspect the entire observation input sequence X. This means that the CRF model avoids the independence assumption between observations and allows nonlocal dependencies between state and observation [18]. Moreover, this model has no label bias problem in contrast with the MEMM. However, the CRF model should assign a label y_t to each time t and do not directly provide a way to estimate the conditional probability of a class label y for an entire sequence X.

The HCRF model shown in Figure 1(d) is a generalized CRF model with hidden states s_t. It incorporates hidden state variables in a discriminative multiclass random field model. By allowing a classification model with hidden states, no a priori segmentation into substructures is needed, and labels at individual observations are optimally combined to form a class conditional estimate. As an augmentation of the CRF, this model can represent long-range dependencies among observations without the label bias problem. The HCRF model was introduced by Quattoni and Gunawardana and has been successfully applied for gesture recognition and phone classification [9, 10]. Due to its advantageous characteristics, however, we believe that the HCRF model can be also successfully applied to vision-based daily activity recognition.

4. Activity Recognition System

We design a system that can recognize high-level daily activities based on the 3D body pose data acquired from Microsoft's Kinect API. A high-level daily activity can be regarded as a hierarchical activity structure consisting of multiple subactivities activities that are performed sequentially or iteratively. For example, the activity of *picking up an object on the floor* consists of three subsequent subactivities: *stooping down*, *grasping the object*, and *standing up*, as described in Figure 2.

For the purpose of our research work, we collect the training data of such high-level daily activities to construct the KAD-30 dataset. The KAD-30 dataset consists of 10 activities in total: opening a lid, drinking water, tying shoelaces, stretching, eating cereal, making a phone call, grasping an object on the floor, putting on and taking off a coat, cleaning the floor and writing on a whiteboard. The proposed activity recognition system consists of three steps: feature extraction, model learning, and activity recognition.

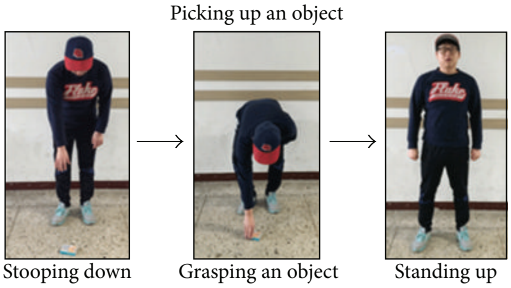


Figure 2: The hierarchical structure in an activity.

4.1. Feature Extraction. In this step, view-invariant features are extracted based on 3D position data from 15 joints of the human body, including the head, neck, and torso, and two sets of joint directional data that correspond to the head and torso. As mentioned before, the set of 3D joint positions are directly provided by Microsoft's Kinect API, which can be estimated from the depth images acquired from the Kinect sensor. However, the 3D position (x, y, z) of each joint provided by Kinect API is represented based on the Cartesian coordinate system of which origin $(0, 0, 0)$ is on the center of the Kinect sensor. Thus, the 3D position data of a joint can be easily changed if at least either the Kinect sensor or the target object changes its position. This means that the 3D joint coordinates of joints directly acquired from Kinect API are very sensitive to Kinect's view variance, and so they are not proper features used to distinguish daily human activities robustly under various environmental conditions. Figure 3 illustrates the view variance problem. As shown in the figure, if Kinect's view is changed, the corresponding position value of the same elbow joint captured by the Kinect sensor will be also changed. In order to meet the view variance problem and get more informative features, our system extracts joint angles from the subject's skeleton model and then performs the feature transformation to get three different types of features regarding motion, structure, and hand positions.

While performing one of the daily activities, each joint of the performer moves according to a specific pattern over time. These temporal patterns of joint movement may be effectively captured by using motion features. In addition, daily activities are considered to be performed through multiple interactions between distinct joints. For example, grasping an object on the floor is mainly accomplished through interaction between the joints of the knee and the hand. We try to capture these spatial patterns through structure features. A lot of human daily activities include hand movement. Unlike other animals, humans use their hands very much to work in daily life. For example, consider when drinking water and opening the lid of a container. Hand position features, which represent the position of both hands relative to the head and the torso, can help distinguish human daily activities using hands.

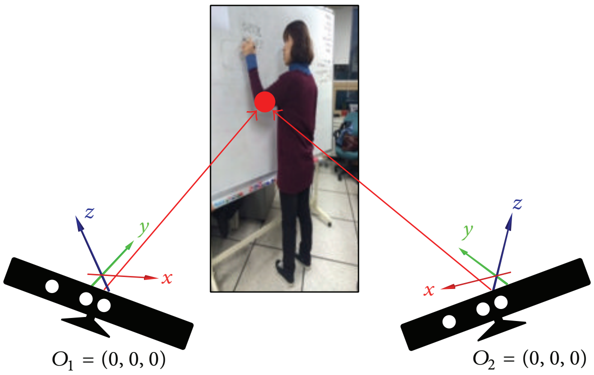


Figure 3: The view variance problem.

Figure 4 illustrates the process of extracting the motion and the structure features. As shown in Figure 4, the 3D Cartesian coordinates of the form (x, y, z) is first transformed into 2D spherical coordinates of the form (θ, ϕ) for each joint, where θ is the polar angle and ϕ is the azimuthal angle of the joint. The following equation shows how to compute the polar θ and the azimuthal angles ϕ from the corresponding 3D joint coordinates (x, y, z). In the equation, r is the radial distance, which is the Euclidean distance from the origin to the joint. In our work, the radial distance r is omitted and only the polar θ and the azimuthal angles ϕ are used to extract features through subsequent processes:

$$\theta = \cos^{-1}\frac{z}{r}, \qquad \phi = \tan^{-1}\frac{y}{x}. \tag{1}$$

From the transformed 2D spherical coordinates $a_{t,n}$ of each joint n, motion features $m_{t,n}$ and structure features $s_{t,n}$ are calculated through the following equations. Below, t and n refer to the frame and joint indexes, respectively:

$$\text{motion } (m_{t,n}) = a_{t,n} - a_{t-1,n}, \qquad \text{structure } (s_{t,n}) = a_{t,n} - a_{t,k}. \tag{2}$$

The motion features $m_{t,n}$ of joint n are obtained from the tth input frame by computing the difference between the current $a_{t,n}$ and the previous position $a_{t-1,n}$ of the joint n. Hence the motion features $m_{t,n}$ represent the positional change of each joint n from the $(t-1)$th frame to the tth frame. On the other hand, the structure features $s_{t,n}$ of joint n are extracted from the tth input frame by computing the difference between the current position $a_{t,n}$ of the joint n and the current position $a_{t,k}$ of the other joint k. Here assume that the joint n is, for example, the center of the head, the joint k can be one of the other joints, such as the neck or the torso. Hence the structure features $s_{t,n}$ represent the relative position of the joint n based on the other joint k at the tth frame. It is assumed that the position $a_{t,n}$ of each joint n at frame t has already been transformed into 2D spherical coordinates $(\theta_{t,n}, \phi_{t,n})$ in the aforementioned way.

Figure 5 describes the process of extracting the hand position features. The position features of each hand are obtained by computing its relative positions with respect to both the head and the torso. For example, while the relative position features $h_{t,\text{left,head}}$ of the left hand with respect to the head are computed through (3), its relative position features $h_{t,\text{left,torso}}$ with respect to the torso are calculated through (5). Similarly, the relative position features $h_{t,\text{right,head}}$ and $h_{t,\text{right,torso}}$ of the right hand are computed through (4) and (6), respectively. In the equations, $j_{t,\text{left\_hand}}$, $j_{t,\text{right\_hand}}$, $j_{t,\text{head}}$, and $j_{t,\text{torso}}$ represent the 3D position vector of the left hand, the right hand, the head, and the torso, respectively. On the other hand, $o_{t,\text{head}}$ and $o_{t,\text{torso}}$ are the 3×3 orientation matrix of the head and the torso, respectively:

$$\text{relative\_left\_hand\_position\_wrt\_head } (h_{t,\text{left,head}}) = (j_{t,\text{left\_hand}} - j_{t,\text{head}}) * o_{t,\text{head}}, \tag{3}$$

$$\text{relative\_right\_hand\_position\_wrt\_head } (h_{t,\text{right,head}}) = (j_{t,\text{right\_hand}} - j_{t,\text{head}}) * o_{t,\text{head}}, \tag{4}$$

$$\text{relative\_left\_hand\_position\_wrt\_torso } (h_{t,\text{left,torso}}) = (j_{t,\text{left\_hand}} - j_{t,\text{torso}}) * o_{t,\text{torso}}, \tag{5}$$

$$\text{relative\_rightt\_hand\_position\_wrt\_torso } (h_{t,\text{right,torso}}) = (j_{t,\text{right\_hand}} - j_{t,\text{torso}}) * o_{t,\text{torso}}. \tag{6}$$

In general, the higher the number of feature vector dimensions, the higher the computational complexity required for model learning and activity recognition. The feature vectors acquired from the feature extraction process have 252 dimensions. Vector quantization is executed by applying k-mean clustering to the high dimensional feature vectors to increase the efficiency of model learning and activity recognition. Through vector quantization, each high dimensional feature vector is replaced into an integer index indicating the cluster the feature vector belongs to. As a result, one-dimensional integer type time-series data is generated while performing an activity. Here, because the length of the time-series data is determined by performing time per activity, a different length per activity is generated. The subsequent processes of the proposed activity recognition system, modeling learning, and activity recognition use these time-series feature data of each activity for the purpose of model training and testing.

4.2. Model Learning. As mentioned before, many high-level daily activities can be regarded as having a hierarchical structure, where multiple subactivities are performed sequentially or iteratively. Our system utilizes the hidden state conditional random field (HCRF) model to represent effectively the hierarchical nature of such activities. In order to recognize a number of activities with a single trained model, our system uses a multiclass HCRF model. A state variable in this HCRF model represents a subactivity belonging to a high-level activity and it is assumed to be hidden. Therefore, there is no

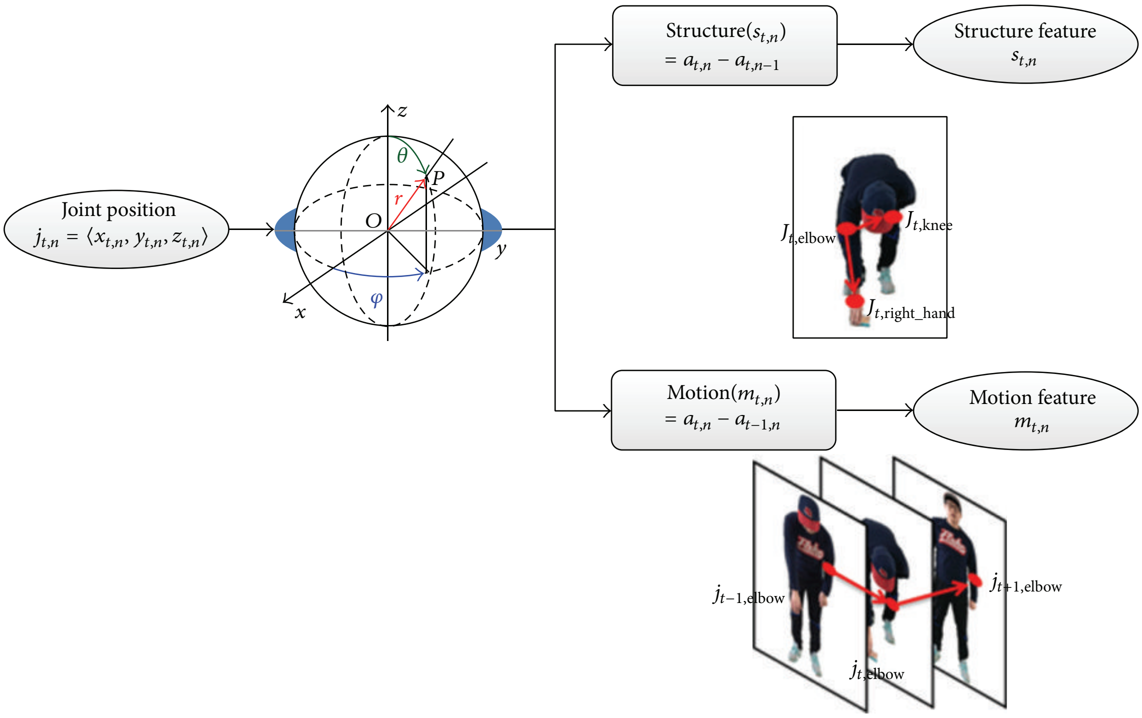


Figure 4: Extracting the motion and structure features.

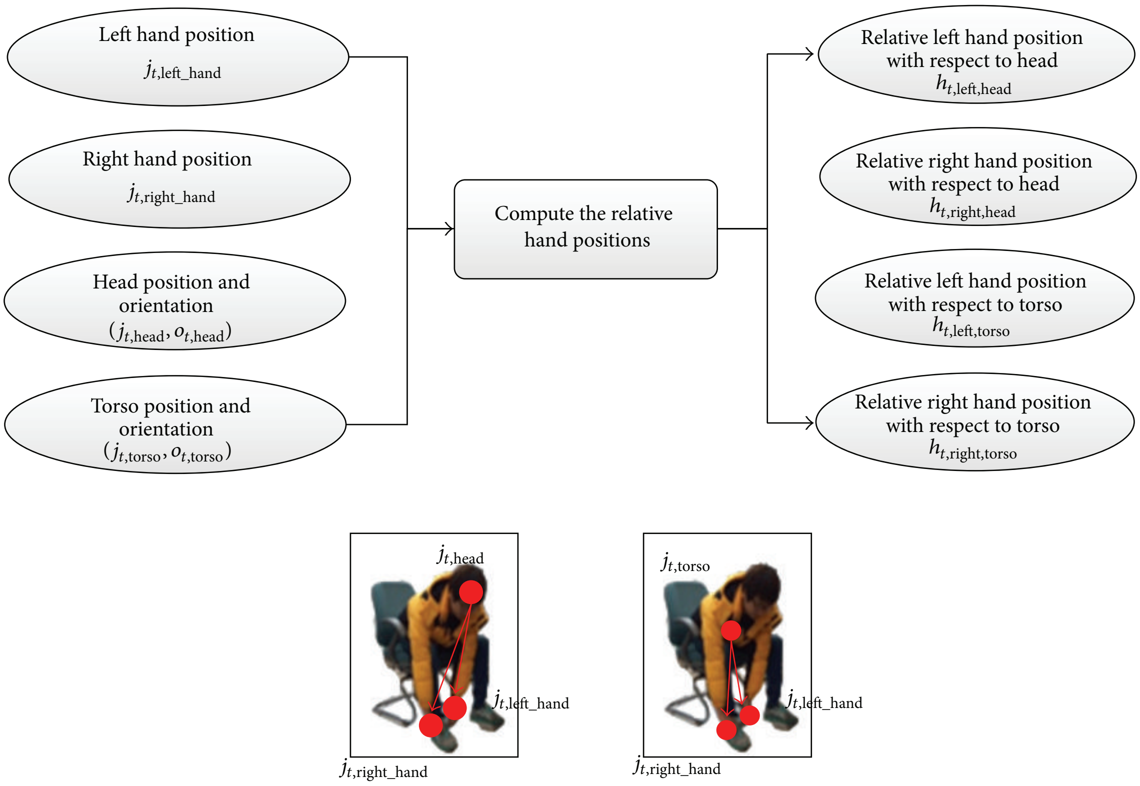

Left hand position
$j_{t,\text{left\_hand}}$
Right hand position
$j_{t,\text{right\_hand}}$
Head position and orientation
$(j_{t,\text{head}}, o_{t,\text{head}})$
Torso position and orientation
$(j_{t,\text{torso}}, o_{t,\text{torso}})$
Compute the relative hand positions
Relative left hand position with respect to head
$h_{t,\text{left,head}}$
Relative right hand position with respect to head
$h_{t,\text{right,head}}$
Relative left hand position with respect to torso
$h_{t,\text{left,torso}}$
Relative right hand position with respect to torso
$h_{t,\text{right,torso}}$
$j_{t,\text{head}}$
$j_{t,\text{left\_hand}}$
$j_{t,\text{right\_hand}}$
$j_{t,\text{torso}}$
$j_{t,\text{left\_hand}}$
$j_{t,\text{right\_hand}}$

Figure 5: Extracting the hand position features.

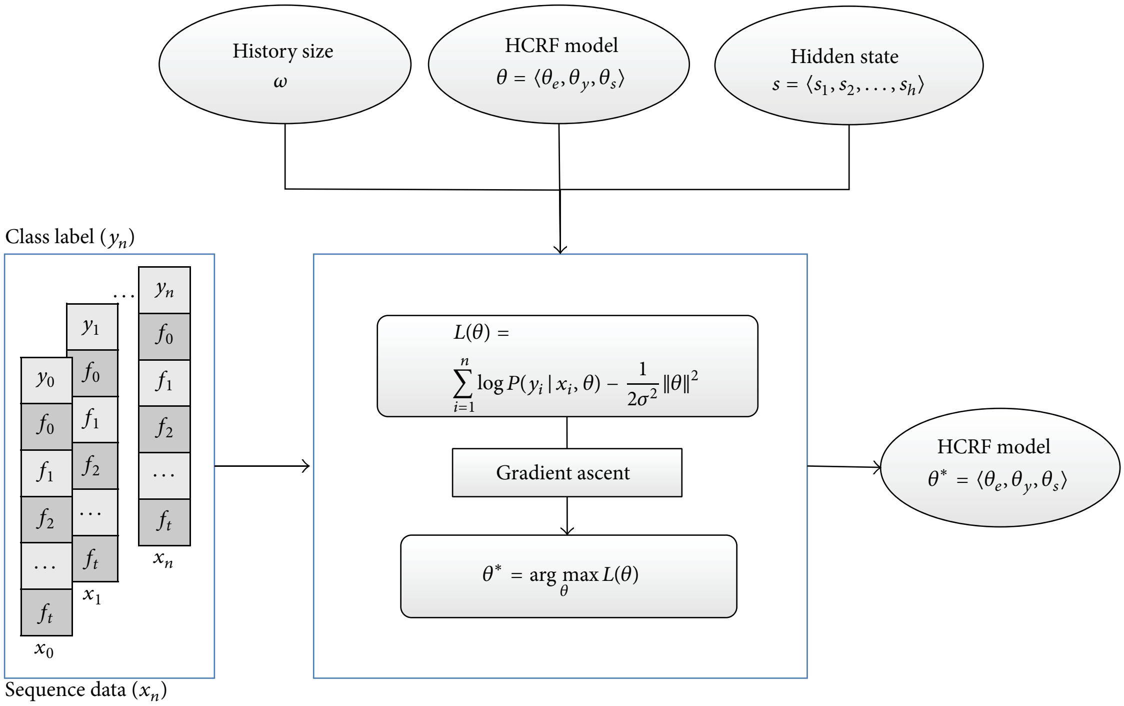

History size
ω
HCRF model
$\theta = \langle \theta_e, \theta_y, \theta_s \rangle$
Hidden state
$s = \langle s_1, s_2, \ldots, s_h \rangle$
Class label (y_n)
y_0
y_1
y_n
f_0
f_1
f_2
...
f_t
x_0
x_1
x_n
Sequence data (x_n)
$L(\theta) = \sum_{i=1}^{n} \log P(y_i \mid x_i, \theta) - \frac{1}{2\sigma^2} \|\theta\|^2$
Gradient ascent
$\theta^* = \arg\max_{\theta} L(\theta)$
HCRF model
$\theta^* = \langle \theta_e, \theta_y, \theta_s \rangle$

Figure 6: Learning parameters of the HCRF model.

need to designate a label for each subactivity in the training data.

Figure 6 shows the process to learn the optimized parameters $\theta^* = \langle \theta_e, \theta_y, \theta_s \rangle$ of the HCRF model. The parameter vector θ^* is made up of three different components: θ_e, θ_y, and θ_s. θ_s refers to the parameters corresponding to state s_j. Similarly, θ_y stands for the parameters corresponding to class y and state s_j. θ_e refers to the parameters corresponding to class y and the pair of states s_j and s_k. In order to learn the optimized parameters θ^* from the initial parameters θ, the training data of the form (x_i, y_i) are used, where x_i is an observation sequence and y_i is the label of activity class.

In the model learning process, the optimized parameters θ^* are searched to maximize the objective function $L(\theta)$ using the training dataset. The first term of the objective function $L(\theta)$ includes the conditional probability $P(y_i \mid x_i, \theta)$. The conditional probability $P(y \mid x, \theta)$ of a class label y given the observation x is defined as in the following equation:

$$P(y \mid x, \theta) = \sum_{s} P(y, s \mid x, \theta) = \frac{\sum_s e^{\Psi(y,s,x;\theta)}}{\sum_{y' \in Y, s \in S^m} e^{\Psi(y',s,x;\theta)}}. \quad (7)$$

The objective function $L(\theta)$ depends on the potential function $\Psi(y, s, x; \theta)$, parameterized by θ, which measures the compatibility among a label, a set of observations and a configuration of the hidden states. Using the gradient ascent method, the optimized parameters θ^* are found to maximize the objective function $L(\theta)$, as in the following equation:

$$\theta^* = \arg\max_{\theta} L(\theta). \quad (8)$$

The number of hidden states h and the size of history ω are determined in advance in order to train the HCRF model. In our system, the number of hidden states of the HCRF model is set to 7, considering the complexity of the target activities. The history size, which determines dependency range, is set to 1. As the optimization function to adjust the weight of feature vectors in the HCRF model, Limited-Memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) is used.

4.3. Activity Recognition. In the activity recognition step, the conditional probability of each activity, $P(y \mid x, \omega, \theta^*)$, is calculated using the trained HCRF model θ^* and the test sequence data x. And then the test data x is recognized as the activity y^* with the highest conditional probability, as in the following equation:

$$y^* = \arg\max_{y \in Y} P(y \mid x, \omega, \theta^*). \quad (9)$$

5. Performance Evaluation

Based on the design explained before, our activity recognition system was implemented using C++ and MATLAB on Windows 7. Several experiments were conducted to evaluate the performance of our proposed activity recognition system. In the experiments, two different datasets are used: the KAD-30 dataset from Kyonggi University and the CAD-60 dataset from Cornell University. Figure 7 shows 10 common daily activities included in the KAD-30 dataset. The activities in the KAD-30 dataset are *opening a lid*, *drinking water*, *tying shoelaces*, *stretching*, *eating cereal*, *making a phone call*, *picking up an object on the floor*, *putting*

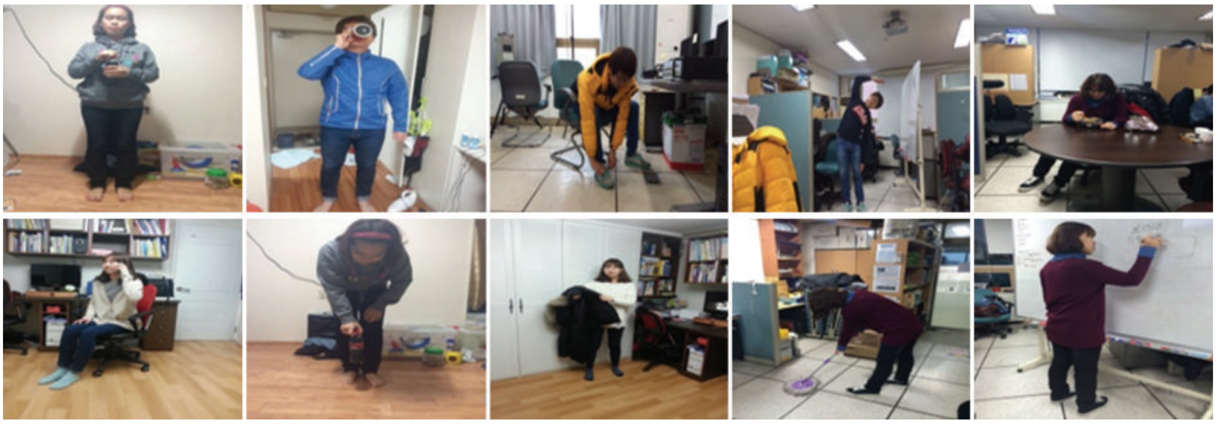


Figure 7: Activities included in the KAD-30 dataset.

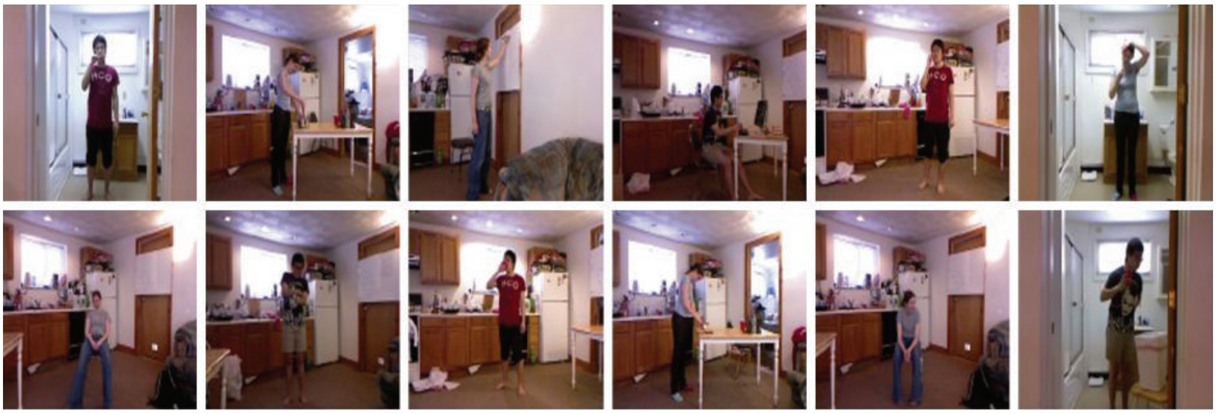


Figure 8: Activities included in the CAD-60 dataset.

on and taking off a coat, *wiping the floor*, and *writing on a whiteboard*. To collect the KAD-30 dataset, 3 different subjects performed 10 different activities ten times in front of the Kinect sensor. 3D body pose data for each activity were recorded for 30 to 40 seconds at 30 frames/second speed.

Figure 8 shows 12 daily human activities in the CAD-60 dataset provided by Cornell University. The activities included in the CAD-60 dataset are *brushing teeth*, *cooking (stirring)*, *writing on a whiteboard*, *working on computer*, *talking on the phone*, *wearing contact lens*, *relaxing on couch*, *opening pill container*, *drinking water*, *cooking (chopping)*, *talking on couch*, and *rinsing the mouth*.

To analyze the performance of our activity recognition system, three different experiments were conducted using the KAD-30 and CAD-60 datasets. In the first experiment, we compared the recognition performance of two different HCRF models: one-versus-all HCRF model and multiclass HCRF model. A one-versus-all HCRF model is able to distinguish only one activity from others. In order to recognize N different activities, a total of N one-versus-all HCRF models need to be learned. On the other hands, the single multiclass HCRF model can be learned to recognize N different activities. In addition, we conducted the experiment with different sizes of history ω to analyze the effect of long-range dependency by setting $\omega = 0$ for one model and $\omega = 1$ for the other.

Table 1: Performance comparisons between two different HCRF models.

HCRF models	Datasets	
	KAD-30	CAD-60
HCRF (one-versus-all) $\omega = 0$	86.33	86.11
HCRF (one-versus-all) $\omega = 1$	90.67	88.27
HCRF (multiclass) $\omega = 0$	91.67	90.23
HCRF (multiclass) $\omega = 1$	92.33	92.18

Table 1 summarizes results of the experiment to compare the recognition performance between the one-versus-all HCRF model and the multiclass HCRF model. The multiclass HCRF model performs better than the one-versus-all HCRF model. The performance of HCRF models made a significant improvement when the history size was increased, which indicates that incorporating long-range dependencies was useful.

In the second experiment, we analyzed the recognition performance per activity of the multiclass HCRF model. For this experiment, we set the history size ω of the multiclass HCRF model to 1. Figure 9 shows two confusion matrices

	Opening a lid	Drinking water	Tying shoelaces	Stretching	Eating cereal	Making a phone call	Grasping an object on floor	Putting on and taking off a coat	Wiping the floor	Writing on whiteboard
Opening a lid	0.93	0.03								0.03
Drinking water	0.03	0.90		0.06						
Tying shoelaces			0.93			0.07				
Stretching				0.96			0.04			
Eating cereal		0.03			0.90	0.06				
Making a phone call	0.03					0.96				
Grasping an object on floor		0.03		0.06			0.90			
Putting on and taking off a coat								0.96		0.03
Wiping the floor							0.1		0.90	
Writing on whiteboard	0.09			0.03						0.86

(a) KAD-30 dataset

	Brushing teeth	Cooking (chopping)	Cooking (stirring)	Drinking water	Relaxing on couch	Opening pill container	Rinsing mouth	Talking on couch	Talking on the phone	Wearing contact lenses	Working on computer	Writing on whiteboard
Brushing teeth	0.96	0.04										
Cooking (chopping)		0.96					0.04					
Cooking (stirring)	0.04		0.96									
Drinking water		0.03		0.94				0.03				
Opening pill container			0.03		0.88	0.03				0.06		
Relaxing on couch						0.96	3.7					
Rinsing mouth	0.04						0.89				0.06	
Talking on couch				0.04				0.89		0.04		0.03
Talking on the phone			0.11						0.85	0.4		
Wearing contact lenses			0.06		0.11		0.04			0.89		
Working on computer								0.04			0.96	
Writing on whiteboard							0.03		0.03			0.94

(b) CAD-60 dataset

Figure 9: Confusion matrix for each dataset.

for the KAD-30 and CAD-60 datasets as results of the experiment. In the case of the KAD-30 dataset, the activity of *writing on the whiteboard* showed the lowest recognition accuracy. This was because the hands of the target subject were often hidden by his/her torso while writing on the board. For the CAD-60 dataset, *opening a pill container* and *wearing contact lens* activities showed lower recognition accuracies than other activities. This was due to insufficient available information as these activities took a shorter time to perform than the others.

In the third experiment, we compared the recognition performance among three different probabilistic graphical models: HMM, CRF, and multiclass HCRF. Due to their inherent assumptions and structures, these models have different power of expression. Therefore, we expect that the activity recognition with different models will result in different performances. Table 2 summarizes the results of the experiment to compare the recognition performance among three different probabilistic graphical models. In this experiment, our multiclass HCRF model with the history size set to one ($\omega = 1$) performs better than the HMM, the CRF, and even the multiclass HCRF model with the history size set to zero ($\omega = 0$). The HMM performed better than the CRF model for both the KAD-30 and the CAD-60 datasets. In this experiment, hidden state models such as HMM and HCRF perform better than nonhidden state models like CRF. This result implies that hidden state models are very effective to learn the hierarchical structure of high-level human activities. We also found that the CRF and the multiclass HCRF models made some improvements when the history size was increased. This result indicates the useful effect of long-range dependencies in the CRF and the HCRF models.

Table 2: Performance comparisons among three different graphical models.

Learning models	Datasets	
	KAD-30	CAD-60
HMMs	90.67	90.84
CRFs $\omega = 0$	86.33	86.88
CRFs $\omega = 1$	88.00	87.86
HCRF (multiclass) $\omega = 0$	91.67	90.23
HCRF (multiclass) $\omega = 1$	92.33	92.18

6. Conclusions

In this paper, we proposed a daily activity recognition system that applies the multiclass HCRF model to Kinect sensor data. The HCRF model is used to represent the hierarchical structure of high-level daily activities in effect. In addition, the proposed system extracts three kinds of view-invariant features from 3D joint coordinates provided by Kinect API. These features represent various characteristics of high-level daily activities. These characteristics include the movement pattern of each joint over time, the structural relationship between two different joints at an instant time, and the relative positions of both hands. Through experiments using the KAD-30 dataset from Kyonggi University and the CAD-60 dataset from Cornell University, the high recognition performance of the proposed system was verified.

In the future, our research highlights would be focused on the following points. On the one hand, we will optimize our system so as to further improve the performance. On the other hand, our system will be extended for many useful applications such as home healthcare, human robot interaction (HRI), and other context-aware services.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This work was supported by the GRRC Program of Gyeonggi province.

References

[1] N. Howard and E. Cambria, “Intention awareness: improving upon situation awareness in human-centric environments,” *Human-Centric Computing and Information Sciences*, vol. 3, article 9, 2013.

[2] J. McNaull, J. C. Augusto, M. Muvenna, and P. McCuuagh, “Flexible context aware interface for ambient assisted living,” *Human-Centric Computing and Information Sciences*, vol. 4, article 1, 2014.

[3] S.-H. Yoon and J. Min, “An intelligent automatic early detection system of forest fire smoke signatures using Gaussian mixture model,” *Journal of Information Processing Systems*, vol. 9, no. 4, pp. 621–632, 2013.

[4] X. Yang, G. Peng, Z. Cai, and K. Zeng, “Occluded and low resolution face detection with hierarchical deformable model,” *Journal of Convergence*, vol. 4, no. 2, pp. 11–14, 2013.

[5] H. T. Manh and G. Lee, “Small object segmentation based on visual saliency in natural images,” *Journal of Information Processing Systems*, vol. 9, no. 4, pp. 592–601, 2013.

[6] K. Goswami, G. Hong, and B. Kim, “A novel mesh-based moving object detection technique in video sequence,” *The Journal of Convergence*, vol. 4, no. 3, pp. 20–24, 2013.

[7] K. Chorianopoulos, “Collective intelligence within web video,” *Human-Centric Computing and Information Sciences*, vol. 3, article 10, 2013.

[8] H. Cho and M. Choi, “Personal mobile album/diary application development,” *The Journal of Convergence*, vol. 5, no. 1, pp. 32–37, 2014.

[9] S. B. Wang, A. Quattoni, L. P. Morency, D. Demirdjian, and T. Darrell, “Hidden conditional random fields for gesture recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR ’06)*, vol. 2, pp. 1521–1527, June 2006.

[10] Q.-B. Gao and S.-L. Sun, “Trajectory-based human activity recognition using hidden conditional random fields,” in *Proceedings of the International Conference on Machine Learning and Cybernetics (ICMLC ’12)*, pp. 1091–1097, July 2012.

[11] A. Jalal, M. Z. Uddin, and T.-S. Kim, “Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home,” *IEEE Transactions on Consumer Electronics*, vol. 58, no. 3, pp. 863–871, 2012.

[12] M. Z. Uddin, N. D. Thang, J. T. Kim, and T.-S. Kim, “Human activity recognition using body joint-angle features and hidden Markov model,” *ETRI Journal*, vol. 33, no. 4, pp. 569–579, 2011.

[13] L. Xia, C. C. Chen, and J. K. Aggarwal, “View invariant human action recognition using histograms of 3D joints,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW ’12)*, pp. 20–27, Providence, RI, USA, June 2012.

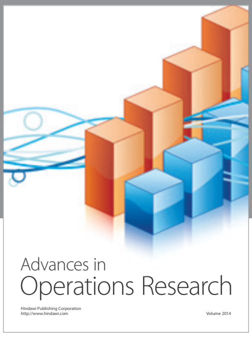
[14] J. Sung, C. Ponce, B. Selman, and A. Saxena, “Human activity detection from RGB-D images,” in *Proceedings of the AAAI Workshop on Plan, Activity, and Intent Recognition (PAIR ’11)*, 2011.

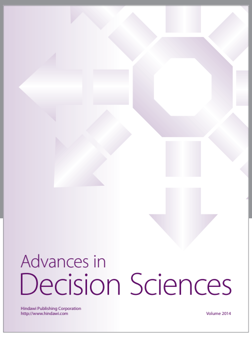
[15] C. Zhang and Y. Tian, “RGB-D camera based daily living activity recognition,” *Journal of Computer Vision and Image Processing*, vol. 2, no. 4, 2012.

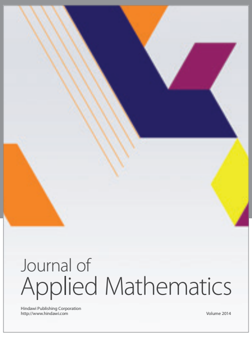
[16] W.-H. Ong, P. Leon, and T. Koseki, “Investigation of feature extraction for unsupervised learning in human activity detection,” *Bulletin of Networking, Computing, Systems, and Software*, vol. 2, no. 1, pp. 30–35, 2013.

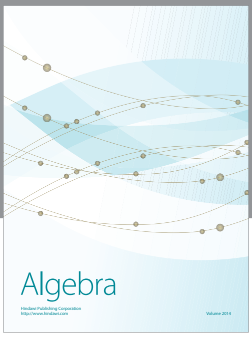
[17] K. Salim, B. Hafida, and R. S. Ahmed, “Probabilistic models for local patterns analysis,” *Journal of Information Processing Systems*, vol. 10, no. 1, pp. 145–161, 2014.

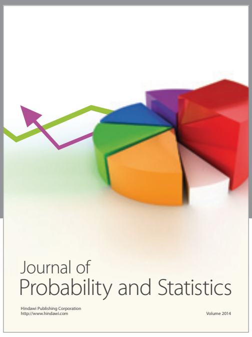
[18] D. L. Vail, M. M. Veloso, and J. D. Lafferty, “Conditional random fields for activity recognition,” in *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS ’07)*, ACM, May 2007.


Advances in
Operations Research

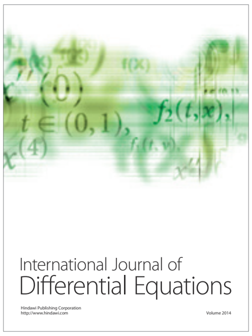

Advances in
Decision Sciences


Journal of
Applied Mathematics

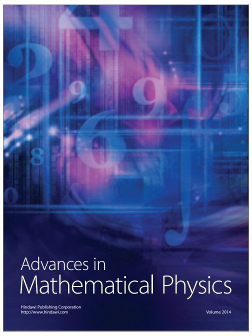

Algebra

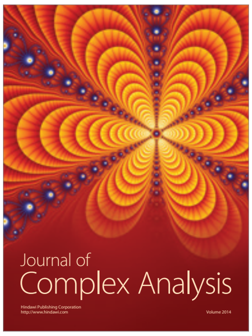

Journal of
Probability and Statistics

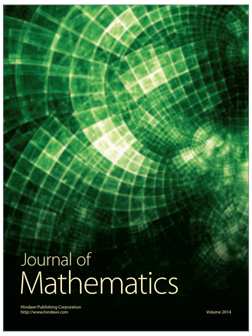
The Scientific
World Journal

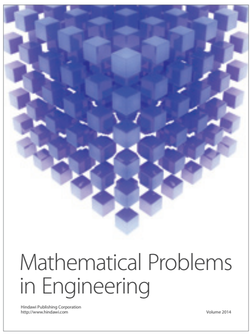

International Journal of
Differential Equations

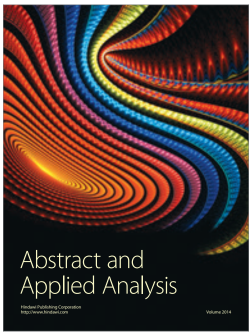
International Journal of
Combinatorics

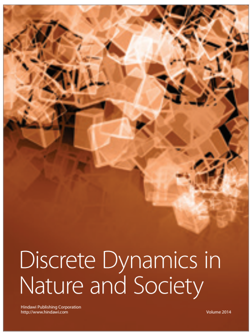

Advances in
Mathematical Physics

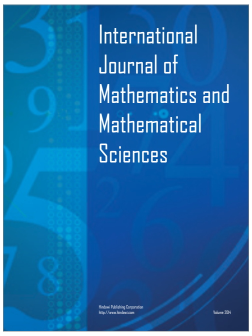

Journal of
Complex Analysis


Journal of
Mathematics


Mathematical Problems
in Engineering

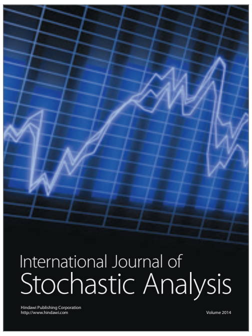

Abstract and
Applied Analysis


Discrete Dynamics in
Nature and Society


International
Journal of
Mathematics and
Mathematical
Sciences

Journal of
Discrete Mathematics

Journal of
Function Spaces


International Journal of
Stochastic Analysis

Journal of
Optimization