Hindawi
Scientific Programming
Volume 2020, Article ID 6349342, 14 pages
https://doi.org/10.1155/2020/6349342

Research Article

Partial Observer Decision Process Model for Crane-Robot Action

Asif Khan,[1] Jian Ping Li,[1] Amin ul Haq,[1] Shah Nazir,[2] Naeem Ahmad,[3] Naushad Varish,[4] Asad Malik,[5] and Sarosh H. Patel[6]

[1] *School of Computer Science and Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu 611731, China*
[2] *Department of Computer Science, University of Sawabi, Swabi, KPK, Pakistan*
[3] *School of Computer Applications, Madanapalle Institute of Technology and Science, Madanapalle, India*
[4] *Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, KL University, Guntur, India*
[5] *School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China*
[6] *Interdisciplinary Robotics, Intelligent Sensing & Control (RISC) Lab, Department of Computer Science & Engineering, School of Engineering University of Bridgeport, Bridgeport, CT, USA*

Correspondence should be addressed to Asif Khan; asifkhan@uestc.edu.cn and Jian Ping Li; jpli2222@uestc.edu.cn

Received 1 December 2019; Accepted 27 December 2019; Published 28 February 2020

Guest Editor: Rahman Ali

The most common use of robots is to effectively decrease the human's effort with desirable output. In the human-robot interaction, it is essential for both parties to predict subsequent actions based on their present actions so as to well complete the cooperative work. A lot of effort has been devoted in order to attain cooperative work between human and robot precisely. In case of decision making , it is observed from the previous studies that short-term or midterm forecasting have long time horizon to adjust and react. To address this problem, we suggested a new vision-based interaction model. The suggested model reduces the error amplification problem by applying the prior inputs through their features, which are repossessed by a deep belief network (DBN) though Boltzmann machine (BM) mechanism. Additionally, we present a mechanism to decide the possible outcome (accept or reject). The said mechanism evaluates the model on several datasets. Hence, the systems would be able to capture the related information using the motion of the objects. And it updates this information for verification, tracking, acquisition, and extractions of images in order to adapt the situation. Furthermore, we have suggested an intelligent purifier filter (IPF) and learning algorithm based on vision theories in order to make the proposed approach stronger. Experiments show the higher performance of the proposed model compared to the state-of-the-art methods.

1. Introduction

Environmental perception and object recognition is an important part of the image processing. It can be widely used in robot visual perception, video surveillance, exception handling, intelligent early warning and rapid retrieval and efficient image storage, camera, and other fields. Humans can perceive the complex scenes easily and respond to get the location and type of the target object correctly, but currently this is a challenging problem for robot visual understanding.

Human eyes also have best capability to capture where neurons help in filtering the scenes. Human motion forecasting is the ability to predict subsequent motion series according to a given sequence of motions. By observing the motion behavior of the target object, the motion features are extracted and then motion forecasting is finally realized. Until now, processing of software as pirated or not pirated still becomes a challenging task [1]. Human beings can realize such forecasting through observation, which embodies the more intelligent reasoning ability of human beings (Figure 1). In some sensitive scenarios where the object is completely unknown (encrypted) from the observer and that object has to be identified in the encrypted form, ideas need to be added in the future work to improve the observer's capability in the encrypted domain [2, 3].

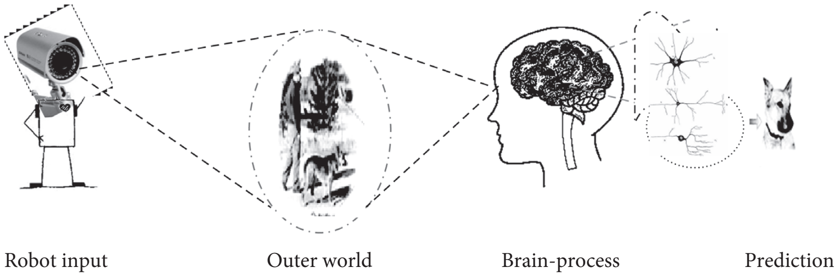


Figure 1: Robot Vs human real-world observation.

The human-robot or robot-robot interactions ability becomes particularly crucial for observing the motion of the object. In this context, an object may vary in orientation and scale or may even be partially obstructed. But, this is not always impairing our ability to recognize it. Due to the lack of this ideal platform, judging a complex scene, obtaining the location, and targeting the object accurately are complex tasks for machines or robots as compared with humans. Computer-assisted predictive system plays an important role to assist any observer recognition [4]. To assist any observer, selected features might be redundant variables which must be handled [5]. Selecting the most appropriate components is crucial for the success of the entire machine. However, decisions regarding software component reusability are often made in an ad hoc manner, which ultimately results in schedule delay and lowers the entire quality system [6].

Recognition by the robot vision is a tough and challenging problem to predict a significant part of the complex, unstructured, and arbitrary scenes; it is also very difficult to balance and place the output of the algorithm and the effect of recognition for already known targets. Visual scenes interact with each other in various topography combinations, i.e., the arrangement of the physical characteristics of a region, and adaptive system design is difficult to enhance understanding of the impact of natural scenes in complex environments.

Therefore, the development of the natural environment, image processing, and computer vision is focused on visual perception and faces enormous challenges. The visual perception system is highly nonlinear dynamic system level neural information collection. For storage and intelligible process, visual attention structure plays an important role in the visual perception. Software for decision making like birthmark is a unique quality to detect software theft [7]. Local visually interpreted information and available computing resources are concentrated on the most essential evidence that makes the visual perception possible in real-time, which can be customized to the dynamic perception of the real world.

Every living thing has patterns of action as per their nature, but for machines there is need to program them to work accordingly. From our home to big industries, there is lot of application of robotics that can be found like vacuum cleaner, self-driving vehicles, and different types of industrial robots. In such types of robots, the working of brain and vision is very similar to human beings' brain and eye control.

A lot of effort is being made in the research and development sector throughout the world to find solutions for this problem [8]. Vision capability towards perception and sensing is a real phenomenon for understanding mobility and manipulation of real world's random situation. In many situations, it is difficult to get sufficient images of an object, which makes the object recognition and the identity authentication difficult. Small sample with high dimension problem is a hot project recently. In a conventional object database, the number of images is limited.

The objectives of this research are encouraged by the neural network cognitive intelligence and the adaptive nature scene recognition technology, which can enhance the understanding of natural scenes, targeting the object, and solving the diversity, randomness, complication of natural scenes and other problems, making the real-time visual system highly flexible. It is to provide a stable foundation for the practical application of mining. Natural complex environment with complex scenes diversification shown in Figure 2 demonstrated how to overcome the lack of randomness in the visual processing system.

For example, the Biological Vision Model (BVM) is devoted to providing a new technological approach on behalf of merging new cognitive visual futures with inspired nerve cells cognitive intelligence cortex, which try to relate with real-world object recognition. To perceive arbitrary natural scene from complex environment perception and sensing in robotic mobility and manipulation on unstructured random natural scene understanding is a challenging problem in visual imaging and processing [9].

Neural network is a map of "neuron like" nodes; in this paper, we are taking neural network (NN) as just an example, committed to making a contribution a new technical concept for scene comprehension and acknowledgment by reorganizing new visual intellectual characteristics into scene expression, which can be very essential and provide robot vision with perceptual intelligence. This approach not only let the system proceed but also provide learning in natural scene with complex environmental perception and understanding. Through the study of perception ability of the natural scene image from complex environment, robot vision is enhanced with the integration of cognitive visual feature and the scene expression [10].

Our contributions summarized as follows:

(i) We enhance the efficiency of capturing, representing the target features of visual images and improving the features representation of the natural environment, so that the system can intelligently observe the unorganized nature scenes.

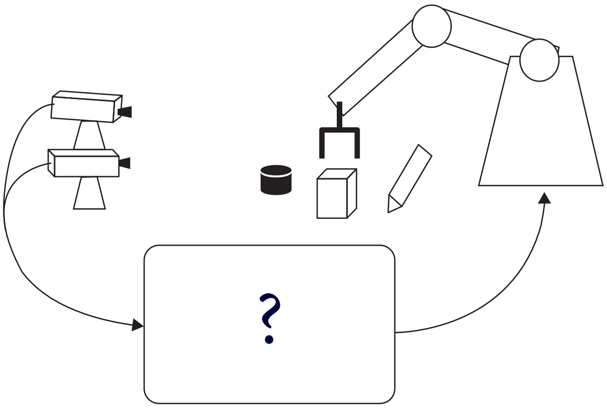


Figure 2: Real-time visual system.

(ii) We proposed a model that can go through essential, generally measured capacity skill for intelligent approach of vision-based information retrieval system, analyzing and refining as a breakthrough to provide better intelligence to the visual information.

(iii) Our proposed model inherits a new intelligent purifier filter processing scheme, that is, upgradation of bio-inspired image processing.

(iv) The proposed model is essentially inspired by complex BM (Boltzmann Machine) mechanism that is scene prediction for visual information processing, which is expert in obtaining better perception for decision performance with deep belief network. We provide considerable empirical observations on the chosen datasets to support obtained results.

The remainder of this paper is organized as follows. In Section 2, we outline the related works with cause of motivation for our work. Section 3 covers the proposed partial observer decision process model which is further described with two subsections: first is to obtain the possible perception for making next step decision with deep belief network, and second is to learn decision for further action by its filter analysis that is included with learning algorithm. In Section 4, the remarkable performance is demonstrated by the experimental simulation and their outcomes. Finally, in Section 5, we conclude our proposed analysis with future accepts.

2. Related Work

Studies have shown that [11] the factors affecting visual attention from two aspects, i.e., top-down prior knowledge and input signal, make the sensor stimulus from bottom to up. Among them, the top-down prior knowledge and applications are highly correlated, which is very tough for modeling analysis. Therefore, there are lot of sensor stimulations only for the Bottom-up visual attention model. The paradigm of bottom-up visual attention can be classified into two categories [12]. One is to use the eye tracker eye to glance at the image's location and use statistical methods to make the eye zone appear longer and as a significant area of human interest. Another category is defined by multichannel multiscale analysis of the input image, statistically significant interest on the extent of each pixel in image depending on the extraction distribution.

First visual attention model based on significant distribution map had been proposed by Koch and Ullman [13]. Before that, there have been many visual attention models based on significant distribution in [13–17]. But there is no model work with a compatible system for complex domain knowledge like human eyes. The same does not involve the human eye gaze input image and gaze time which statically show the number of repetitive tests that are pleasant to human beings like us. There is many field space where target detection [11, 18, 19], video compression and coding [20, 21], image analysis [22, 23] and scene understanding [24]. These models will be applied. And other fields can use the limited memory computing resources to process the input video image or the region of most interest of human vision.

Therefore, without decreasing the intervention efficiency of the concept, the system not only reduces the overhead space but also increases system performance in many ways, such as the effect of processing individual vision needs is more, a stronger noise robustness has increased stability in complex backgrounds, and so on [11, 18, 19, 22–25]. These models, moreover, need to calculate multiscale and multichannel features of the Gaussian pyramid input image and calculate these considerable sample dividend payments into a globally significant distribution, using the Winner-Take-All (WTA) mechanism to select the most significant area independently [24].

The entire process requires a large number of intermediate results that can be stored and has a larger amount of computation, making it even more difficult to implement the limited computing resources in embedded systems. Biological science experiments confirmed that in the primate temporal cortex of the brain, nerve cell activity and animal identification of objects are closely linked [25]. When contrasting with the generic image model stored in the brain, the reorganization of the particular object can be understood. The researchers therefore conclude that a viable approach is to simulate the visual cortex structure in order to construct the object recognition.

Related to the earliest primate visual system model is the neocognitron model [26], which is based on self-organizing feed-forward neural networks. British Wallis and Rolls Experimental Psychology Department of the University of Oxford promoted the constant target identification VisNet primate model [27] and the improved version called Visnet2 [28]. It is a four-level feed-forward, convergence, and competitive nature of the network, where each layer brings together the former cell layer in a small portion of the input field (called filters). With this aggregation law, primate visual cortex cells increase the size of the receptive field characteristics by simulating the advance from junior to senior level. Mel made the SEEMORE model in 1997, and it is also a feed-forward hierarchical structure model which uses the color, shape, and texture combination to achieve visual object recognition. SEEMORE used multiclass combination of features to improve the recognition robustness. Serre et al. in 2005 and 2007 applied HMAX model to object

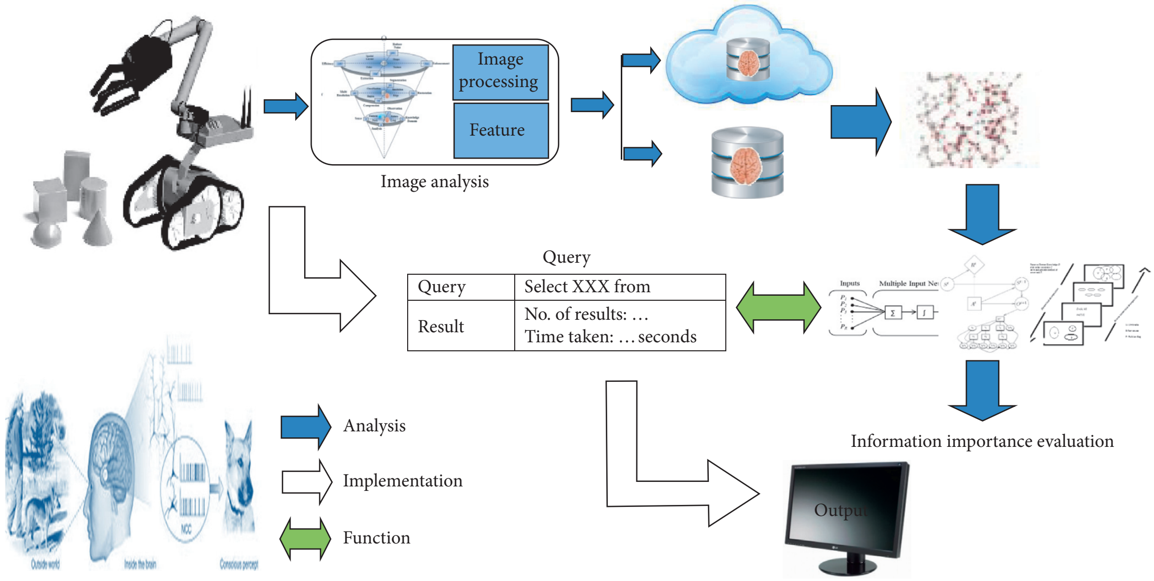


Figure 3: Observer prediction and decision process.

recognition; the improved model constructed high-level simulation of biological visual features. Visual features of the hierarchy template is a operation of matching and merging along the object recognition, where continuous simulation is used for invariant scale, translation, and rotation in the visual cortex. A lot of researchers have made outstanding contributions in this field.

Our proposed model has two aspects. (a) It provides an intelligent platform for integration of features and preprocesses to predict future prediction. For this we analyze Boltzmann machine mechanism [29], whose outcomes go through the second phase of purifier intelligent filter, which is inspired by the biological vision model to purify, segment, and identify the object, which makes the proposed model simple and efficient. (b) Second aspect covers the decision-based model that is incorporated based on accurate perception results, and the partners can cooperate with each other better. This requires the observer to possess the ability to identify and estimate motion sequences [30, 31].

3. Partial Observer Decision Process

To provide vision intelligence for action, the robot needs to go though the learning of the steps for task, while new associated algorithms are put forward to settle range of demanding theoretical problems in visual information processing system. To explore the inherent characteristics that provide new visibility for perception, such as diversity, randomness, and complexity in real-time complex natural environment, where adapting Network perception ability of the natural scene image is improved with the combination of cognitive visual features and scene expression.

The perception hierarchical model outcomes directly incorporate and participate for sensing object decision model. Action model can concurrently remember more than one objective, not only for the common goal of better classification, but also on the texture, non-rigid targets classification. This model is based mainly on visual computing simulation to calculate a cortical action (set of task) network hierarchy. The process model of observation can be easily understood by Figure 3.

To predict information C about dynamic object in complex environment by using visual information and predict about, typical approach form input $(A, \widehat{X})$/output $(B, \widehat{Y})$ relationships as follows,

$$f_0\left(C_{xy}\right) = \begin{cases} A, & \text{for} \quad f_i\left(\widehat{x}, \widehat{y}\right) \leq t, \\ B, & \text{for} \quad f_i\left(\widehat{x}, \widehat{y}\right) > t, \end{cases} \tag{1}$$

where t is the threshold value and f is input/output image functions, respectively. It is ample to implement apparent contrast target and its background. The vision source included in the system makes a complex environment captured to be processed. During this process, the visual feedback is constantly followed to see the template matching for each frame of object's information and forecasting the position extraction dynamically. When the sum squared error between two objects is less than the captured image and BM results in predetermined threshold, then we can say that the object has found the one we were looking for [32].

3.1. *Perception for Decision.*

Boltzmann machine (BM) mechanism is a variant of forecasting outcomes for sensing and perception. BM is a nonlinear generative model for time series that uses an undirected model with binary latent variables, h, connected to a collection of visible variables, v. At each time step t, v and h receive directed connections from the visible variables at the last N time steps, where N is the size of the temporal window considered. The "history" vector or knowledge datasets are concatenated by the data at $t-1, t-2, \ldots, t-N$, which we call $v_{<t}$. In MB, the model

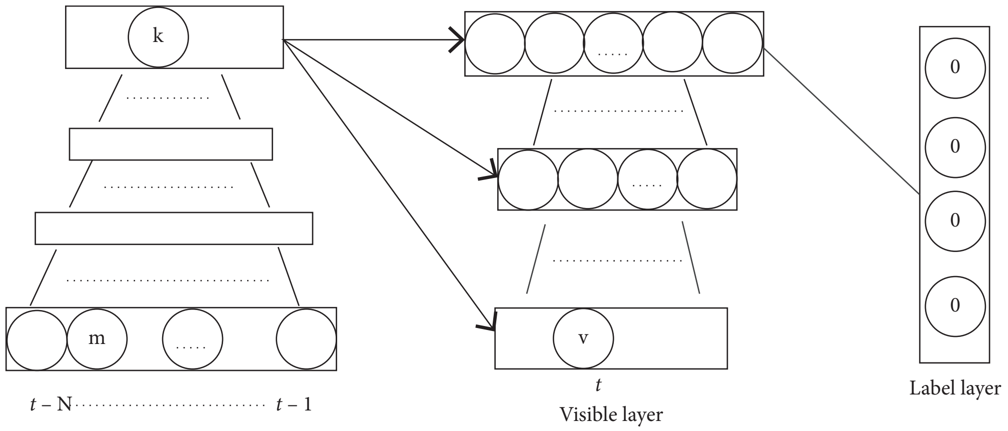


Figure 4: Deep belief network.

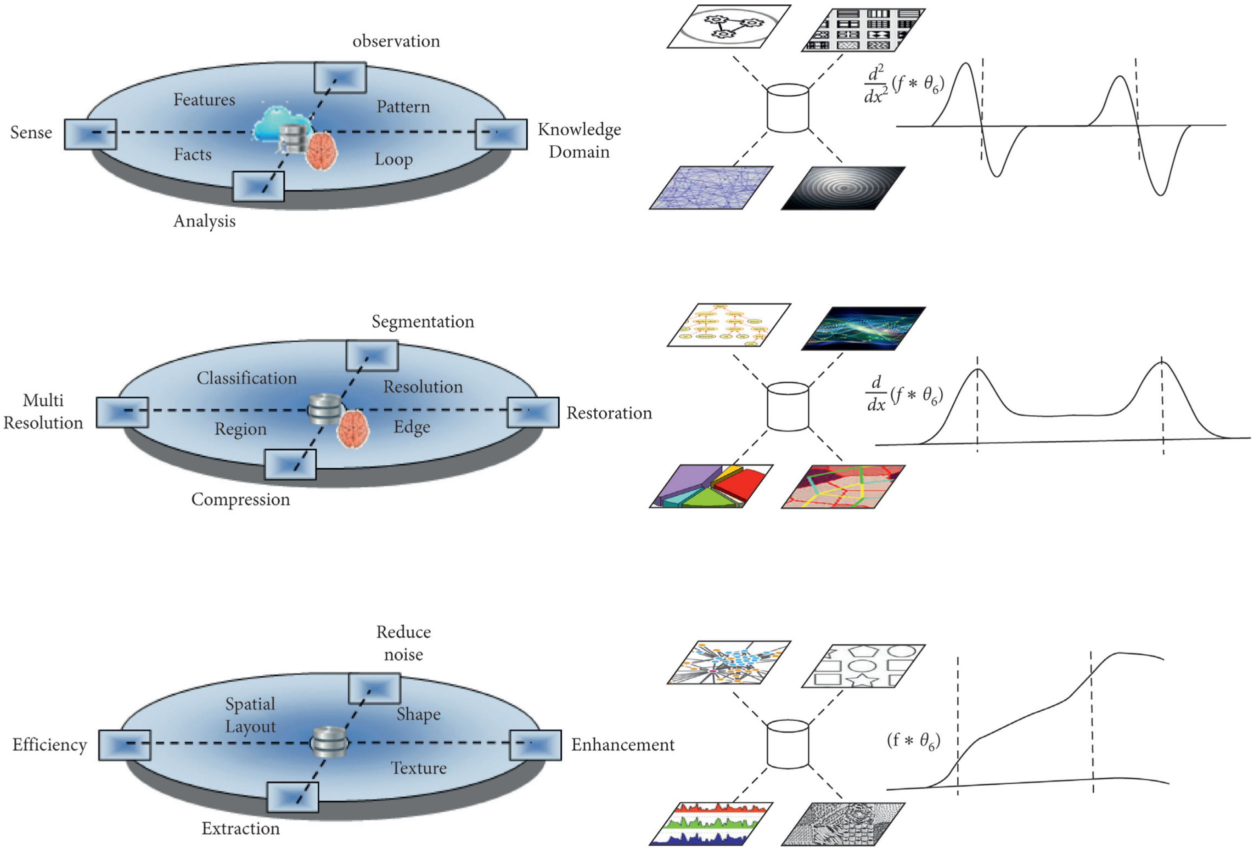


Figure 5: Primely integrated intelligent purifier filter.

defines a joint probability distribution over v_t and h_t, as shown in equation (3) conditioned on $t \geq v$:

$$Z(v_t, h_t \mid v_{<t}) = -\sum_j v_{j,t}\widehat{X}_{j,t} - \sum_i h_{i,t}\widehat{Y}_{i,t} - \rho, \tag{2}$$

where

$$\begin{aligned}\widehat{X}_{j,t} &= bv_j + \sum_k X_{kj} v_{k,<t},\\ \widehat{Y}_{i,t} &= bh_i + \sum_k Y_{ki} v_{k,<t},\end{aligned} \tag{3}$$

where

$$P(v_t, h_t \mid v_{<t}, \theta) = \frac{\exp\left(-Z(v_t, h_t \mid v_{<t}, \theta)\right)}{E(v_{<t})}, \tag{4}$$

where $E(v_{<t})$ is a constant called the partition function and $\widehat{X}_t$ and $\widehat{Y}_t$ are the dynamic biases on time t, which express the input from the past to the visible and hidden units of vision resource (equations (2) and (4)).

Boltzmann machine can also be classified as trained with a generative learning objective, where internal entity

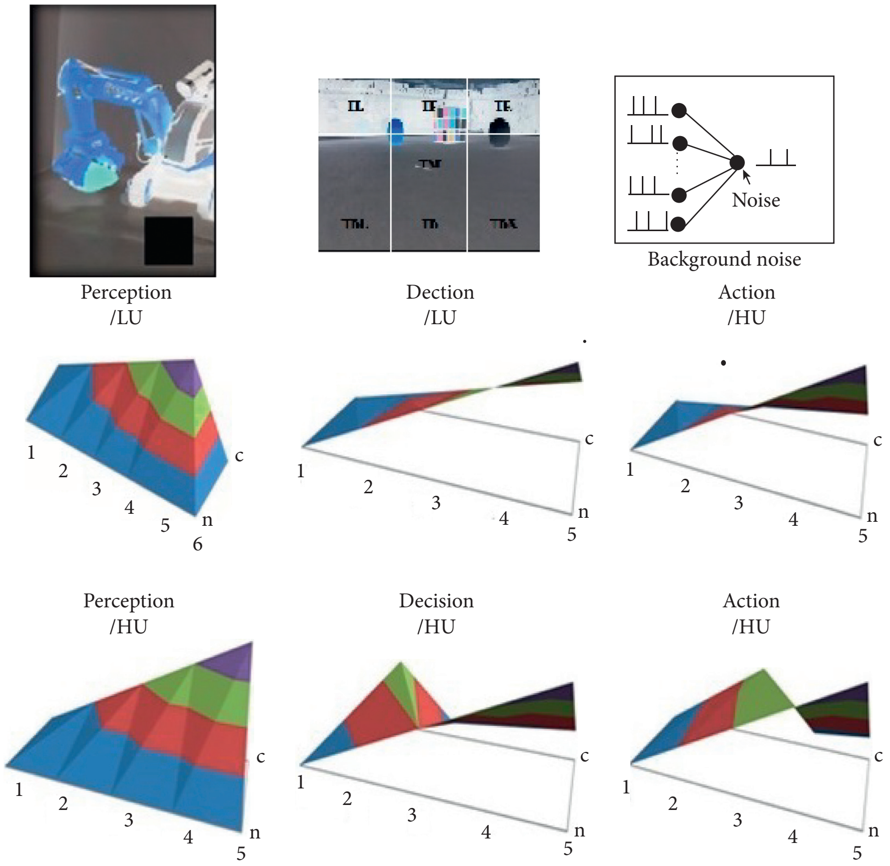


Figure 6: Observer decision process and its partial task analysis.

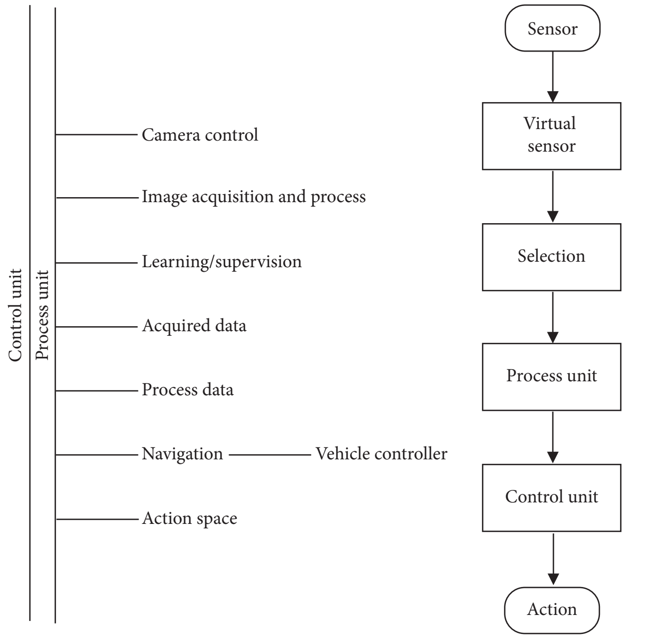


Figure 7: Perception and decision model for next action.

like camera follow this to track, motion, and control. For that learning of the joint distribution $p(v, y)$ of the input vector v and the target class y, and/or a discriminate learning objective and learning of the conditional distribution $p(y|v)$ directly are necessary. It requires no additional training phase for a classifier like the traditional Boltzmann Machine. The energy function of the BM is shown in the following equation:

```
Input: dynamic input coordinates $x_t, y_t$;
Hidden representation of time $t + 1$: $X_{t+1}$;
Class label for the time $t$: $Y_t$;
Output: the prediction on time $t + 1$: $x_{t+1}$;
Step 1: initialization
do
    Calculate the hidden representation on time $t + 1$:
    Calculate the visible representation on time $t + 1$:
    Calculate the class label of the prediction on time $t + 1$:
while $Y_{t+1} == Y_t$;
Adjust the initial input value $x_t$ randomly.
return $Y_{t+1}$;
```

Algorithm 1: Observer parameter learning.

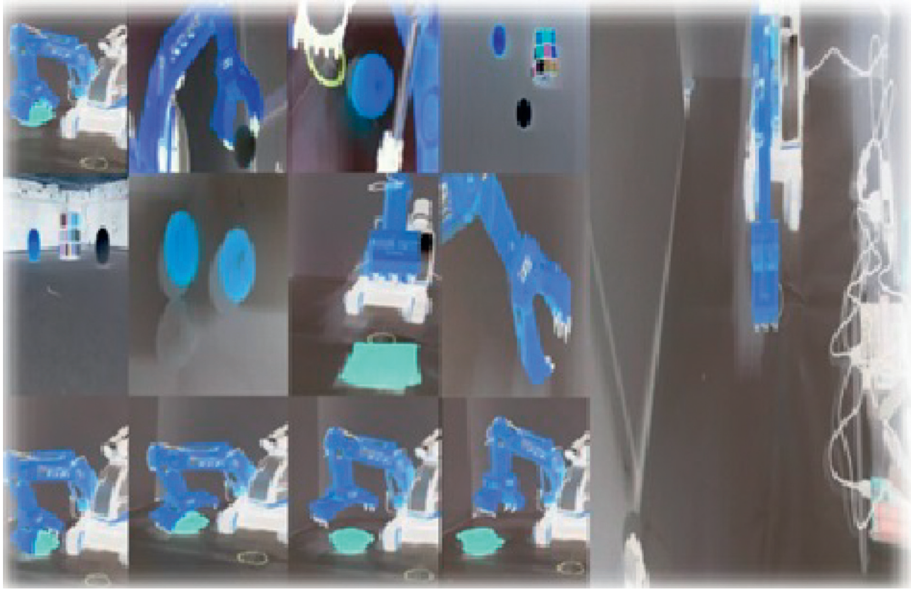


Figure 8: Experimental setup.

Table 1: Forward walk table.

Steps/count (10)	Steps/count (50)
8.5236	31.3214
8.0325	28.8976
7.3178	26.7849
7.4592	24.8934
8.1562	21.2123

$$Z(y, v, h) = -b^T v - c^T h - d^T e_y - h^T \left(Ue_y - Wv\right), \qquad (5)$$

with parameter $\rho = (b, c, d, W, U)$ and where v is the input vector and y is the first process of the class label. To achieve the discriminate objective, the posterior probability in the BM can be inferred from the following equation:

$$P(y|v) = \frac{\exp(-Z(v, y))}{\sum \exp\left(-Z\left(v, y^*\right)\right)}. \qquad (6)$$

The denominator sums over all labels y^* to make $P(y - v)$ a probability distribution. BM can only do the classification task with the independent relation between samples except time series; for the time series, the samples between each other are dependent with each other and can be affected by previous and succeeding samples.

The main purpose of our model, which is primarily based on BM, is to suppress the error amplification problem and prolong the perception length. From the analysis in this section area, we can conclude that the problem that makes BM inefficient during perception primarily arises from two aspects: the first is the previous past result as input data directly and the second is that there is no constraint on the present result. In our work, we should avoid the past prediction result directly being the input data; meanwhile, we should also lower the perception ratio [9].

We retrieve the feature for decision using BN for the last N time steps and discriminate the class label of the

Forward walk table

Steps/count (10)	Steps/count (50)
8.5236	31.3214
8.0325	28.8976
7.3178	26.7849
7.4592	24.8934
8.1562	21.2123

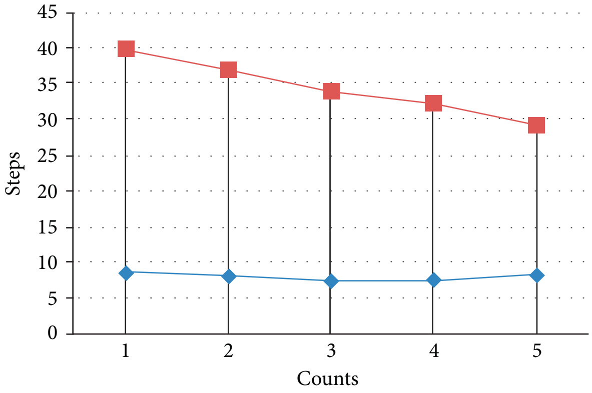


Figure 9: Ratio of walking (forward) table.

Table 2: Backward walk table.

Steps/count (10)	Steps/count (50)
8.8756	36.2543
8.9546	29.6435
7.1289	28.7543
8.1136	26.1235
8.0128	22.8456

forecasting result to decide whether to accept it accordingly. The structure of the model is shown in Figure 3.

3.1.1. Deep Belief Network. After we have trained the model, we can add layers as in a DBN (Figure 4). The previous steps are kept and connected to each hidden layer with an independent weight matrix. The next level will take the previous hidden state vector as the "observed or predicted" data. A two-level model is shown in Figure 4.

3.2. Decision for Action. Intelligent purifier filter (IPF), vision paradigm understanding, and interpretation totally depend on an intelligent relationship between processing of visual real world as inputs and processing output, which make the machine capable to see and understand. Primitive layer and comparative layer integrate all visual processing features and preprocesses by which analysis of image becomes easier. Nowadays different work environments also maintain a huge database of complex world's objects, so by this model, we try to analyze objects on behalf of their preprocess and features as in Figure 5. Models goal layer capable to perform an intelligence of identification, observation with their artificial approach, cause of this possibility second phase of this proposed model reduce complexity of vision for robots to predict accurate for next steps. Every layer consists of different unit type and uses the previous output as the input. (a) The first layer uses the fundamental scale image as the input, and the last layer output is the characteristic value which can be applied to class recognition. Along with the time, field's size increment and complexity become progressive receptively. (b) The complexity of the top visual area is simply built up by the lower-layer steps and has some redundancy. (c) In this proposed model, the purifying filter with the Gaussian pyramid based on the input dynamic objects or real world surrounding calculated brightness, color, direction, and multi-scale characteristics of the channel. It leads to a plenty of calculation and storage of the next process of random sampling. Some others replace across decision for the combinations and normalization with the local extremism method, iterative method, or a prior knowledge method.

DOG filter function can be described as

$$D_o G_s, I_c(I) = G_{\sigma(s)}\left(I - I_c\right)) - G_{s\cdot\sigma(s)\left(I-I_c\right)}, \qquad (7)$$

$$G_{\sigma(s)}(I) = \frac{1}{2\lambda\cdot\sigma(s)^2}\cdot e\left(\frac{1}{2\cdot\sigma(s)}\right). \qquad (8)$$

As per equation (8), it is clear that in 2D for Gaussian function with equation (8) variance $\sigma(\&)$ that depends on the scale as S_1 is the position for center position I_c of filter with photoreceptor.

Then, cell activation is computed as dot product as shown in the following equation:

Backward walk table

Steps/count (10)	Steps/count (50)
8.8756	36.2543
8.9546	29.6435
7.1289	28.7543
8.1136	26.1235
8.0128	22.8456

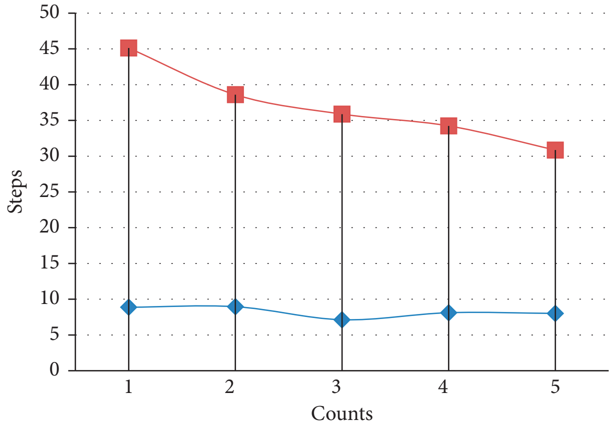


Figure 10: Ratio of walking (backward) table.

$$\langle I, \phi_1 \rangle = x = \sum_{i=R} I(l).\phi_1(l). \quad (9)$$

Here, ϕ_1 is the filter weight, I is the neuron from the respective region R for intensity $I(l)$. So, after the MAX operation, the response of a complex unit for C_1 is

$$r = \max\left(x_j\right), \quad j = 1, 2, 3, \ldots, m. \quad (10)$$

So, precisely timed action potential through intensity is expressed as

$$T_i = f(s_i) = T_{\max} - \log(\beta s_i + 1). \quad (11)$$

That show for one pixel S_1, with scaly factor β maximum time of esocdy windows is $T_{\max}$.

There exists each cell count with different oscillations (subthreshold membrane oscillations). So, it is described as

$$\mathrm{OSC}_i = (\cos(\omega T + \phi)). \quad (12)$$

The number of cycle W and initial phase for I pixel are

$$\phi_1 = \phi_0 + (i-1) \cdot \mathrm{d}\phi. \quad (13)$$

After converting the intensity value, equation (13) into a line action, the algorithmic operation is implemented as all steps of performed relational information.

So, retrieval for learning perform correlation measure is adopted to measure the similar degree between desired (d) and actual o/p. So, matrix epochs are

$$C = \frac{\vec{V}_d \cdot \vec{V}_a}{\left|\vec{V}_d\right| \cdot \left|\vec{V}_a\right|}. \quad (14)$$

After information to learning, we make a decision by correlation C between the desired o/p and actual o/p, so target pattern considers as much closed to C.

The authors believe that due to probabilistic prediction and sensing technique, the ideal (Figure 6) solution provides basic means for extending the capacity of hardware system beyond the boundaries provided by currently used observation process methods (as shown in Figure 7). In our opinion, the application of introduced methods (in particular, the new DBN, purifier filter, and obtained dataset representation of Algorithm 1) leads to effective improvement outcomes, at least for the case of using the following.

4. Experimental Setup and Result

Following, motion, and control are essential for a settled verbalized robot like crane with a portable device like mobile sensor to arrange the objects accordingly with arms and onboard camera, with visible device that is able to position controllable world stage. So these genuine exploratory methodologies consolidate with an algorithmic coding of MATLAB [32] and that signal activity executed as control and process (Figure 8). The action-3D database is a type of motion-action behavior dataset. This dataset was captured by a strength camera. There are 23 types of behavior in the dataset, namely, move-backward, move-forward, move-jump, move-up, move-down, high arm, horizontal arm, hammer, hand catch, throw, draw, circle, hand two hand move, side move, back move, role, left side move, right side move serve, pick up, and throw. Each movement was repeated two times by 10 subjects; thus, there were 30 sequences of each action in the dataset, and there were 600

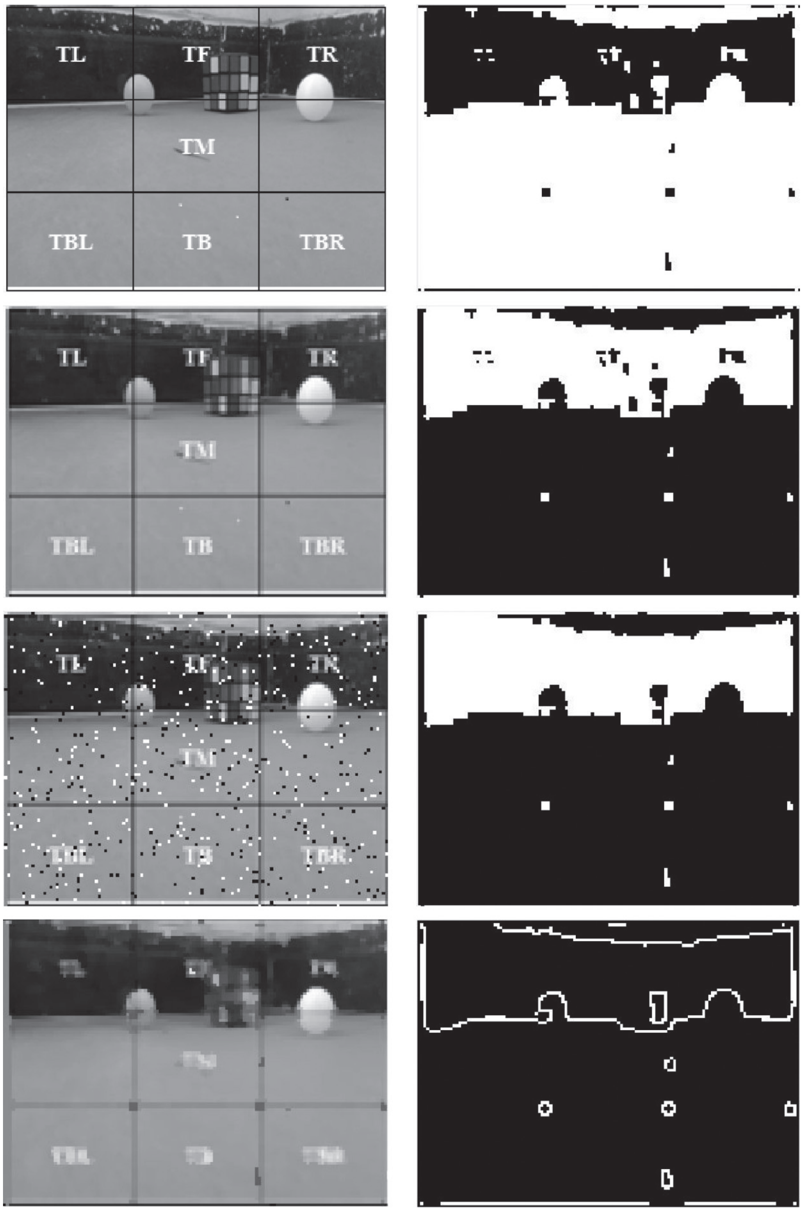


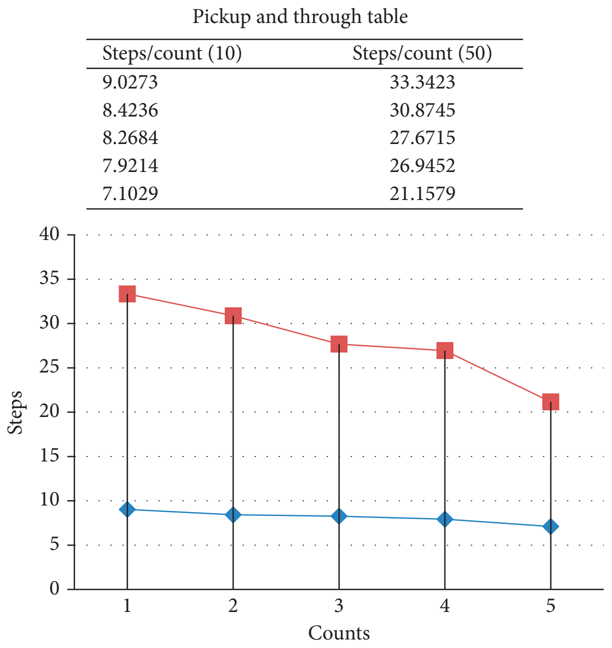
Figure 11: Partial observation analysis and outcomes after LR/HU.

Table 3: Pickup and through table.

Steps/count (10)	Steps/count (50)
9.0273	33.3423
8.4236	30.8745
8.2684	27.6715
7.9214	26.9452
7.1029	21.1579

sequences in total. The sampling frequency is 15 times per second, and the resolution of each frame is 640∗480.

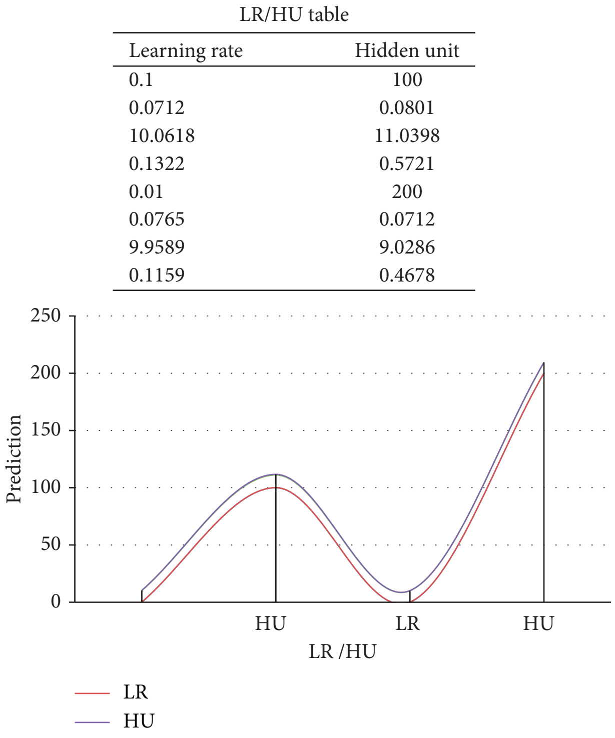
4.1. Simulation Result. We validated the performance of our model on the dynamic observer dataset in IPF. The dynamic observer dataset is retrieved from a video camera. There are two categories in this dataset (as shown in Table 1 and Figure 9). One is "forward walking," where the performer points to somewhere with nothing in their hand, and the other is "backward walking," where the performer steps try to holds the object (as shown in Table 2 and Figure 10). There are a total of 200 time series in the data. We chose 150 series as training data, and the remainder are testing data. Each series contains 150 frames, and each frame is univariate. We represented the whole series in a matrix, and each row stands for a single motion. The preprocessing of data is a very necessary step for good representation of data and machine learning [33]. We show the curve graphs of the two types for output (Figure 11). The left one is the "forward walking" class, and the right one is the "backward walking" class. Then, we incorporate the whole system to grasp the target object and replace at particular destination in final action (as



Pickup and through table

Steps/count (10)	Steps/count (50)
9.0273	33.3423
8.4236	30.8745
8.2684	27.6715
7.9214	26.9452
7.1029	21.1579

Figure 12: Ratio (pickup and through) table.



LR/HU table

Learning rate	Hidden unit
0.1	100
0.0712	0.0801
10.0618	11.0398
0.1322	0.5721
0.01	200
0.0765	0.0712
9.9589	9.0286
0.1159	0.4678

Figure 13: LR/HU for observer prediction.

shown in Table 3 and Figure 12). We took each single time series as a batch, which means that there were a total of 150 batches when training. We first verified the BM with a shallow structure. We show the results of different hidden unit numbers, different prestep numbers, and different learning rates in Figure 13 and Table 4. Because there is no theory on how to set the hidden unit number, prestep number, and learning rate, we chose the root mean squared error (RMSE), mean absolute percent error (MAPE), and mean relative error (MRE) as the criteria. We found that when the prestep number is 5 and the learning rate is 0.01, after 200 iterations, the hidden unit number is 200, which can obtain relatively small value on RMSE, MAE, and MRE. In addition to the hidden unit number, we conducted several experiments with different prestep numbers, and we found that increasing the prestep number cannot improve the prediction performance much further.

Table 4: LR/HU table.

Learning rate	Hidden unit
0.1	100
0.0712	0.0801
10.0618	11.0398
0.1322	0.5721
0.01	**200**
0.0765	0.0712
9.9589	9.0286
0.1159	0.4678

The values in Table 4 show that when the number of hidden units is 200, relatively small values of the three criteria can be obtained. Therefore, in the following experiments, we set the number of hidden units as 200, the learning rate as 0.01, and the previous step number as 5.

4.2. Result Analysis. Because the original data sets are depth images that have high noise, the image is too vague and has other shortcomings; thus, this paper uses the real-time tracking algorithm to extract the image in 3D joint positions and finally combine the 3D dataset vector. Because the motion of the subjects in the dataset is actually 3D stereo motion, we transform the three-dimensional vector into a two-dimensional vector to express the original motion.

4.2.1. Comparison. Based on the results of the base work experiment (Figures 14 and 15) [32], and according to the proposed model outcomes as (in Figure 11, Table 4, and Figure 13) set the layer 2, the previous input step is 5, the hidden unit numbers are 200 for layer 1 and 100 for layer 2, and the learning rate is 0.01. We trained the model for 500 epochs. We divide the dataset into batches. Each batch contains 100 samples. The parameters are updated after each batch. To depict the affection of proposed model, we randomly chose one sequence from the forward-move-action and input the first 5 frames into our model to generate the following 25 frames, hoping that the model can generate the remaining motions correctly. From analyzing the graph, the first predictions of these models are all very close to the targets.

5. Conclusion

This technique provides a simple and efficient approach for vision-based decision and action in comparison with the conventional or traditional one. However, the performance may be influenced by the limitations of the hardware such as

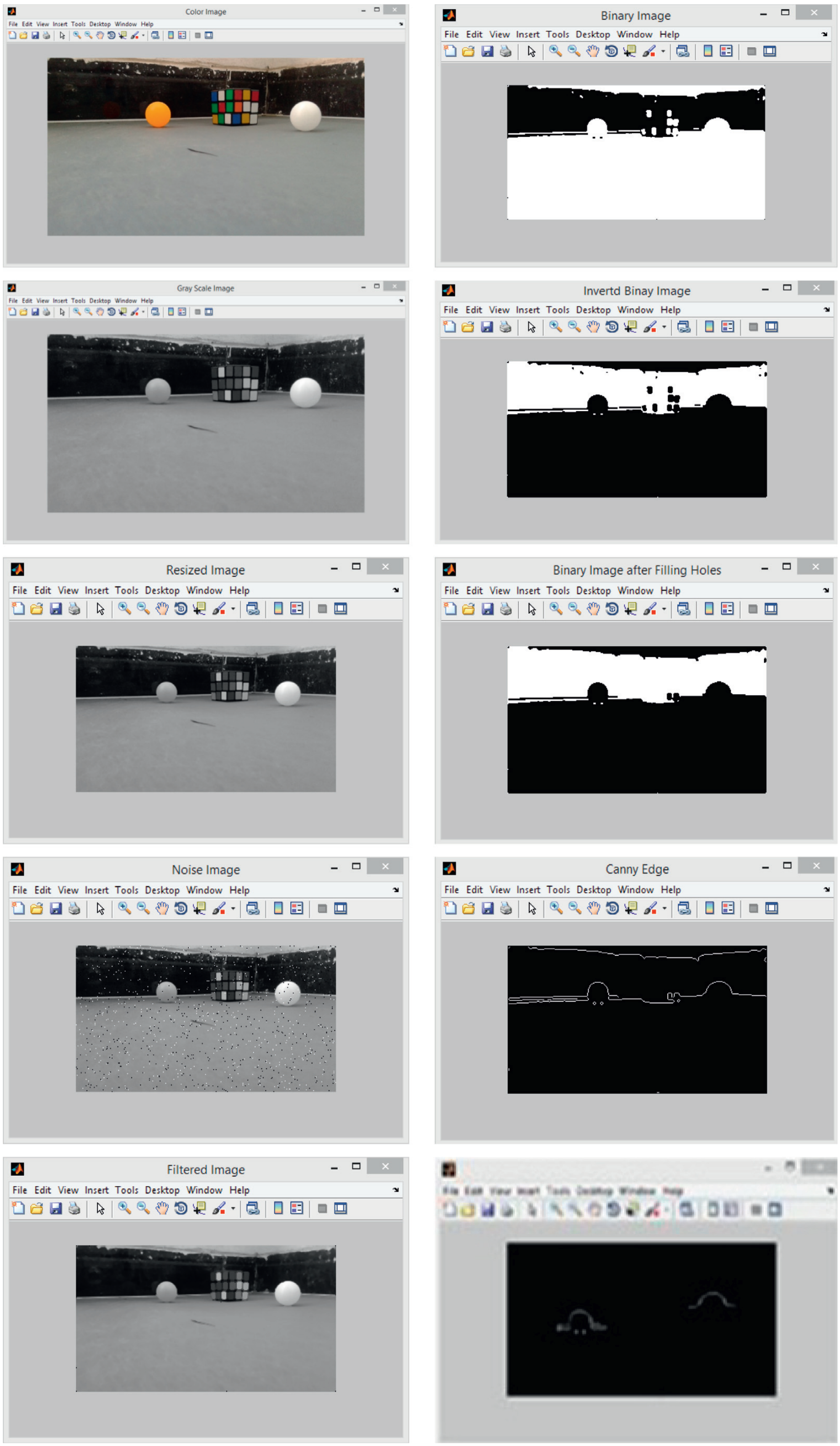


Figure 14: Partial observation analysis and outcomes.

model architecture and decision processing required. Acceptable empirical results have been obtained using the proposed strategy. As per obtained results, the number of preceding inputs and the contemporary decision output considerably have an effect on the performance. The next step in our research will be how to change the number of steps of previous inputs and how many units should be in the hidden layer to produce a high quality result. We will also continue to refine the algorithm to improve prediction and accuracy of practice. We are taking into consideration the speedy action processing and motion estimation as our future work.

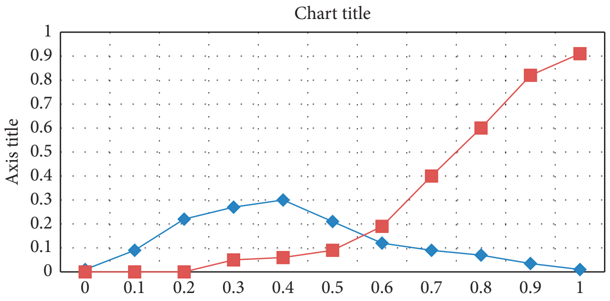


Figure 15: PDF outcome ratio.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant no. 61370073), the National High Technology Research and Development Program of China (grant no. 2007AA01Z423), and the project of the Science and Technology Department of Sichuan Province. All resources used for experiment support were given by Key Laboratory of Wavelet Active Media Technology School of Computer Science, University of Electronic Science and Technology of China (UESTC), No. 2006, Xiyuan Avenue, West Hi-Tech Zone, Chengdu, Sichuan, 611731, P. R. China.

References

[1] S. Nazir, S. Shahzad, and L. S. Riza, “Birthmark-based software classification using rough sets,” *Arabian Journal for Science and Engineering*, vol. 42, no. 2, pp. 859–871, 2017.

[2] A. Malik, H. Wang, H. Wu, and S. M. Abdullahi, “Reversible data hiding with multiple data for multiple users in an encrypted image,” *International Journal of Digital Crime and Forensics*, vol. 11, no. 1, pp. 46–61, 2019.

[3] A. Malik, H. Wang, T. Chen et al., “Reversible data hiding in homomorphically encrypted image using interpolation technique,” *Journal of Information Security and Applications*, vol. 48, Article ID 102374, 2019.

[4] A. U. Haq, J. P. Li, M. H. Memon et al., “Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson’s disease using voice recordings,” *IEEE Access*, vol. 7, pp. 37718–37734, 2019.

[5] A. Ul Haq, J. Li, M. H. Memon, J. Khan, and S. Ud Din, “A novel integrated diagnosis method for breast cancer detection,” *Journal of Intelligent & Fuzzy Systems*, pp. 1–16, 2019.

[6] S. Nazir, S. Anwar, S. A. Khan et al., “Software component selection based on quality criteria using the analytic network process,” *Abstract and Applied Analysis*, vol. 2014, Article ID 535970, 12 pages, 2014.

[7] S. Nazir, S. Shahzad, S. A. Khan, N. Binti Alias, and S. Anwar, “A novel rules based approach for estimating software birthmark,” *The Scientific World Journal*, vol. 2015, Article ID 579390, 8 pages, 2015.

[8] S. Nazir, S. Shahzad, R. B. Atan, and H. Farman, “Estimation of software features based birthmark,” *Cluster Computing*, vol. 21, no. 1, pp. 333–346, 2018.

[9] L. Xia, J. Lv, and D. Liu, “A motion classification model with improved robustness through deformation code integration,” *Neural Computing and Applications*, vol. 31, no. 12, pp. 8519–8532, 2019.

[10] A. Khan, S. Deep, J.-P. Li, K. Kumar, R. A. Shaikh, and F. Hasan, “Vision prehension with cbir for cloud robo,” in *Proceedings of the 2014 11th International Computer Conference on Wavelet Actiev Media Technology and Information Processing (ICCWAMTIP)*, pp. 293–296, IEEE, Chengdu, China, December 2014.

[11] L. Itti and C. Koch, “A saliency-based search mechanism for overt and covert shifts of visual attention,” *Vision Research*, vol. 40, no. 10–12, pp. 1489–1506, 2000.

[12] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau, “A coherent computational approach to model bottom-up visual attention,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 802–817, 2006.

[13] C. Koch and S. Ullman, “Shifts in selective visual attention: towards the underlying neural circuitry,” in *Matters of intelligence*, pp. 115–141, Springer, Berlin, Germany, 1987.

[14] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[15] D. Walther and C. Koch, “Modeling attention to salient proto-objects,” *Neural Networks*, vol. 19, no. 9, pp. 1395–1407, 2006.

[16] T. Kohonen, “A computational model of visual attention,” in *Proceedings of the International Joint Conference on Neural Networks, 2003*, vol. 4, pp. 3238–3243, IEEE, Portland, OR, USA, July 2003.

[17] K. Lee, H. Buxton, and J. Feng, “Cue-guided search: a computational model of selective attention,” *IEEE Transactions on Neural Networks*, vol. 16, no. 4, pp. 910–924, 2005.

[18] J. Han, K. N. Ngan, M. Li, and H.-J. Zhang, “Unsupervised extraction of visual attention objects in color images,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 1, pp. 141–145, 2006.

[19] Z. Chen, J. Han, and K. N. Ngan, “Dynamic bit allocation for multiple video object coding,” *IEEE Transactions on Multimedia*, vol. 8, no. 6, pp. 1117–1124, 2006.

[20] L. Itti, “Automatic foveation for video compression using a neurobiological model of visual attention,” *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1304–1318, 2004.

[21] P. Zhang and R.-S. Wang, “Detecting salient regions based on location shift and extent trace,” *Journal of Software*, vol. 15, no. 6, pp. 891–898, 2004.

[22] C. M. Privitera and L. W. Stark, “Algorithms for defining visual regions-of-interest: comparison with eye fixations,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 9, pp. 970–982, 2000.

[23] C. Siagian and L. Itti, “Rapid biologically-inspired scene classification using features shared with visual attention,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 300–312, 2007.

[24] P. Burt and E. Adelson, “The laplacian pyramid as a compact image code,” *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, 1983.

[25] D. A. Leopold, I. V. Bondar, and M. A. Giese, "Norm-based face encoding by single neurons in the monkey inferotemporal cortex," *Nature*, vol. 442, no. 7102, pp. 572–575, 2006.

[26] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 411–426, 2007.

[27] K. K. Evans and A. Treisman, "Perception of objects in natural scenes: is it really attention free?" *Journal of Experimental Psychology: Human Perception and Performance*, vol. 31, no. 6, pp. 1476–1492, 2005.

[28] M. Carrasco, B. McElree, K. Denisova, and A. M. Giordano, "Speed of visual processing increases with eccentricity," *Nature Neuroscience*, vol. 6, no. 7, pp. 699-700, 2003.

[29] Y. Bengio, "Learning deep architectures for AI," *Foundations and Trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[30] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Advances in neural information processing systems*, pp. 568–576, Montreal, Canada, December 2014.

[31] J. Donahue, L. Anne Hendricks, S. Guadarrama et al., "Long-term recurrent convolutional networks for visual recognition and description," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2625–2634, Boston, MA, USA, June 2015.

[32] A. Khan, J.-P. Li, A. Malik, and M. Yusuf Khan, "Vision-based inceptive integration for robotic control," in *Soft Computing and Signal Processing*, pp. 95–105, Springer, Berlin, Germany, 2019.

[33] A. U. Haq, J. Li, M. H. Memon et al., "Comparative analysis of the classification performance of machine learning classifiers and deep neural network classifier for prediction of Parkinson disease," in *Proceedings of the 2018 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, pp. 101–106, IEEE, Chengdu, China, December 2018.