A proposal model using deep learning model integrated with knowledge graph for monitoring human behavior in forest protection

Van Hai Pham[1], Quoc Hung Nguyen[2], Thanh Trung Le[2], Thi Xuan Dao Nguyen[2], Thi Thuy Kieu Phan[2]
[1]Faculty of Computer Science, School of Information and Communication Technology, Hanoi University of Science and Technology, Hanoi, Vietnam
[2]School of Business Information Technology, University of Economics Ho Chi Minh City (UEH), Ho Chi Minh City, Vietnam

Article Info

Article history:

Received Jul 20, 2021
Revised Sep 14, 2022
Accepted Sep 24, 2022

Keywords:

Deep learning
Forest protection
Human action recognition
Identifying human behavior
Video time-lapse

ABSTRACT

In conventional monitoring of human behavior in forest protection, deep learning approaches can be detected human behavior significantly since thousands of visitors' forest protection is abnormal and normal behaviors coming to national or rural forests. This paper has presented a new approach using a deep learning model integrated with a knowledge graph for the surveillance monitoring system to be activated to confirm human behavior in a real-time video together with its tracking human profile. To confirm the proposed model, the proposed model has been tested with data sets through case studies with real-time video of a forest. The proposed model provides a novel approach using face recognition with its behavioral surveillance of the human profile integrated with the knowledge graph. Experimental results show that the proposed model has demonstrated the model's effectiveness.

Corresponding Author:

Van Hai Pham
Faculty of Computer Science, School of Information and Communication Technology
Hanoi University of Science and Technology, Hanoi, Vietnam
Email: haipv@soict.hust.edu.vn

1. INTRODUCTION

Recently, it is hard for a human to protect forests, indiscriminately cutting down makes forest resources recover and become more and more exhausted, many places where forests can no longer regenerate, the land becomes more reclaimed. The role of forests in environmental forest protection is significant to the world. Identifying human behavior using advanced technology has become an important area of research to create or improve applications that monitor human activity. The behavioral human is a time series of graphs, which is significant for long-term monitoring results. This knowledge graph can be kept track of human actions for human behaviors. Knowledge graphs that represent structural relations among entities such as places, actions, geography nodes, and other attributes of human profiles.

In addition, a knowledge graph has represented an object such as entities, relationships, and semantic descriptions. Computer vision using either analysis or machine learning approaches is automatically detected face and human behavior in real-time. Knowledge graphs represent object types consisting of (i.e. places, geometrics, image coordinates, and locations) and are considered as Neo4j software for making the graphs. After identifying attributes and objects in the graph, the relationships between objects by using geometric and graph attributes. There have been many studies suggesting some solutions to detect bad behavior of humans who may destroy the forest in a forest domain. In this paper, we have proposed a novel approach using a deep learning model integrated with a knowledge graph for the surveillance monitoring system to be activated to confirm human behavior in a real-time video together with its tracking human profile.

The case study of human behavior for forest protection is applied to confirm the proposed model. In the experiments, the proposed model has been tested with data sets through case studies in a real-time video of a forest. Furthermore, the knowledge graph is used to integrate with the proposed deep learning model to make the right decisions person having a normal or abnormal status in forest protection, as shown in all relations of the person profile.

Experimental results show that the proposed model has demonstrated the model's effectiveness. The proposed model provides two types of functions including face recognition with its behavioral surveillance in real-time of a forest. The contribution of this study is: 1) deep learning model with an adaptive prioritization mechanism for the surveillance monitoring system to be activated to confirm human behavior in real-time; and 2) face recognition with its behavioral surveillance in real-time stored in the knowledge graph. By motivating the need for such a system, the proposed system can be performed with surveillance-based contexts such as tracking forest protectors, destroyed tree foresters, forest crimes, and loggerheads. All activities of these profiles can be considered through a face person recognition together with behavioral surveillance in real-time for the forest domain, which is stored in the knowledge graph. Experimental results indicate that the proposed model has been tested for demonstration in this method's effectiveness at 96.38 %.

2. RELATED WORK

Recently, Yuan *et al.* [1], [2] have proposed aerial vehicles (UAVs) with computer vision based on systems for monitoring and detecting forest fires. Sudhakar *et al.* [3] have also used UAV to capture images by using color human recognition and smoke monitoring in the classification of recognition fire. Furthermore, Sudhakar *et al.* [3] have proposed fire detection with forest protection monitoring, using infrared and visual cameras to analyze geographic zones. These techniques are used in forest fire detection using the Voronoi map [4] as monitoring forest fire detection using the Voronoi map and its updated information.

To identify actions in the studies [5], analyzing images collected from human behavior detection cameras [6] these studies apply deep learning algorithms and the studies [7] proposed a method of monitoring the behavior of workers with the framework of vision-based unsafe action detection for behavior monitoring in motion datasets extracted from videos. Zerrouki *et al.* [8] have discussed body structure analysis, tracking, and recognition with good results. Meng [9] has also proposed a taxonomy of 2D approaches, 3D approaches, and recognition, detecting the fall event by adaptive boosting algorithm identifies the human action recognition based on variation in body shape [10]. The studies [9]-[12] have investigated decision intelligence in context-aware systems to provide service provision based on an entity's context, an entity has been defined as "a person, place, or physical or computational object" to track the human profile with a reasoning approach.

Behavior recognition based on intelligent terminals is an emerging research branch of pattern recognition [13], [14]. The acceleration sensor is used to obtain acceleration data information when the user is active, and the data is analyzed to determine the user's behavior category [15], [16]. Norris *et al.* [17] placed the thigh and calve using two accelerometers to obtain the movement information of the human behavior; Fan *et al.* [18] identified common behaviors of the human body in daily life by accelerometers carried in five positions of the human body. Filippeschi *et al.* [19] introduced the accelerometer sensor to the three-axis acceleration information acquired by the front and rear arms of the right hand to realize the recognition of the upper limb movement. Su *et al.* [20] used a single lumbar sensor to obtain gait information, using functional data analysis and a hidden Markov model (HMM) to combine human recognition. Deep learning refers to a learning function model composed of multiple network layers, which is used to extract the characteristics of input data and the abstract features of high-latitude for data classification and combination to obtain more structured results. As a result, to better obtain the characterization of different behaviors, this paper will use the long-term short-term memory model [21], [22] long short-term memory (LSTM), and the deep convolution network model to extract features. Further studies have proposed human action recognition using videos and exporting corresponding tags with outputs of 2D images [23], [24]. The deep learning approaches for human behavior recognition can be considered in this study for the domain of forest protection [25].

3. THE PROPOSED MODEL

3.1. Overview of the proposed model

In this paper, we have developed a proposed model of behavior in forest protection is described as shown in Figure 1, the system consists of 3 main modules of the following:

- Face recognition module.
- Behavior monitoring module.

- Knowledge graph represents a graph database: containing data about the person’s personal information, photos of a person’s face focused on the camera as well as the history of human behavior in the resource fores.

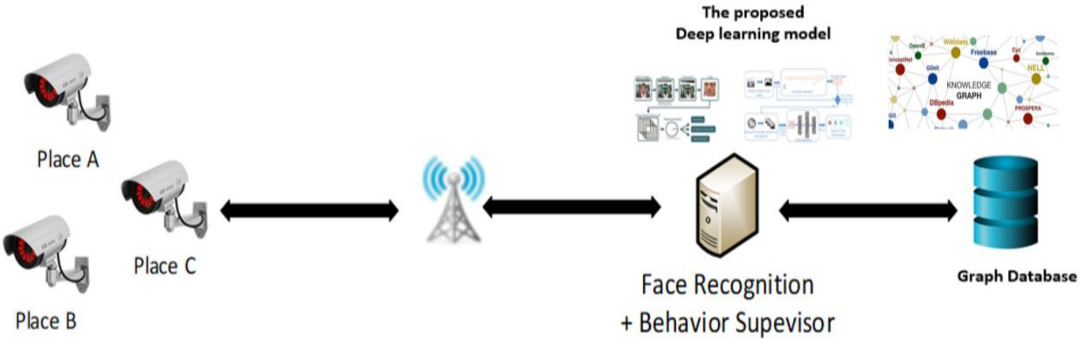

Figure 1. Proposed overall architecture

In the proposed model as shown in Figure 1, a video is essentially a series of consecutive pictures. These pictures grouped in the classification in images of a video are essentially based on the classification in each image of the video. To figure out this problem, two things need to be solved as: 1) image processing and sequence-to-sequence classification; and 2) Implementation of a model that combined both the deep learning model with its graph database by tracking human behaviors as well as profile. The system consists of 3 main modules.

1) Face recognition module:
   a) Model building: the model uses the ResNet101 network structure as the backbone, and the loss function uses the ArcFace model is proven which helps the model converge faster. The model is trained as an identity classification model and will be removed from this layer after the training is completed to get the feature vector every time an image is included.
   b) Architecture faces recognition model: when the face recognition model operates, a model is needed to help locate the face in the image, which is called face detection. Some popular face detection models are Dlib, Haar cascades, and multi-task convolutional neural network (MTCNN). The MTCNN model gives the best results, although the implementation time is longer than the remaining models. Therefore, the face detection model which the author uses is the MTCNN model.
2) Behavior monitoring module: the idea of cameras that can automatically monitor behavior (people and nature) detecting abnormal behavior and early warning for humans has long been rekindled. Many models have been created and achieved certain results. In the framework of this project, the author proposes a combination of convolutional neural network (CNN) cumulative network model and sequence-to-sequence model to solve the problem of human behavior classification in ecological forests.
   a) Similar models: with the born of the CNN model, video classification models have been increasingly improved. The current video classification models are mostly modeling human behavior and activities. The task of these models is to focus on guessing what people are doing in a video. This model may seem similar to the image classification model, but the difference comes from the more difficult level of the video classification problem. In an image classification problem, a machine learning model only has to look at a single picture to make a prediction, for the video model must consider all the frames in that video and one more important thing is that frames are continuous. Its means that the frames in a video follow a chronological order. In the machine learning model, we have a series of problems for similar timed data types, we call the model layers for these problems “time series”.
   b) Idea: a video is essentially a series of consecutive pictures; the classification of a video is essentially based on the classification and processing of each image in that video. To solve this problem, two things need to be solved: image processing and sequence-to-sequence classification and also how to install a model that can combine both models above. The sequence-to-sequence model (recurrent neural network (RNN) and LSTM) solves very well problems where the input is not one but many consecutive data points. Therefore, the above models are often applied to language problems, and chart predictions (stocks). For the image processing model, the convolution network has been proved to be the best model through a series of articles as well as built models. A typical convolution network usually has the first network blocks that are edge detection, followed by

convolution blocks that play a role in shape detection. Following these blocks are fully connected layers that play a role in synthesizing “features” learned from convolution blocks, the output of these layers is the feature vector of the image. Depending on the purpose of the network, this feature vector will be used for classification problems, and face recognition.

c) Building model: the model uses conversion learning techniques, with CNN’s network using InceptionV3 is model trained on ImageNet series. From this CNN network, you will get feature vectors of 2048 dimensions. In addition, a similar model with the backbone is the NasNetLarge network, which was also built to compare the experimental results of the two models. Both InceptionV3 and NasNetLarge models are trained on Google’s ImageNet dataset. The accuracy of InceptionV3 is lower than NasNetLarge (top 1 accuracy 0.779 vs 0.825) but due to its compact structure, InceptionV3 calculation speed is faster than NasNetLarge.

3) The knowledge graph is stored in the graph database: containing data about the person’s personal information, photos of his face as well as the history of his behavior in the resource forest. As shown in Figure 2, to capture video online using cameras, we have defined a series of terms of the following elements: actions, activities, and behaviors. a) actions are descriptions and conscious movements made by humans (e.g. cutting trees and destroying trees); b) activities are combined several actions (e.g. preparing-cutting trees and destroying forests); and c) human behaviors describe how the person performs these activities in real-time.

In the proposed model as shown in Figure 2, a video is essentially a series of consecutive pictures. These pictures grouped in the classification in images of a video are essentially based on the classification in each image of the video. To figure out this problem, two things need to be solved as: 1) image processing and sequence-to-sequence classification; and 2) Implementation of a model that combined both the deep learning model with its graph database by tracking human behaviors as well as activities.

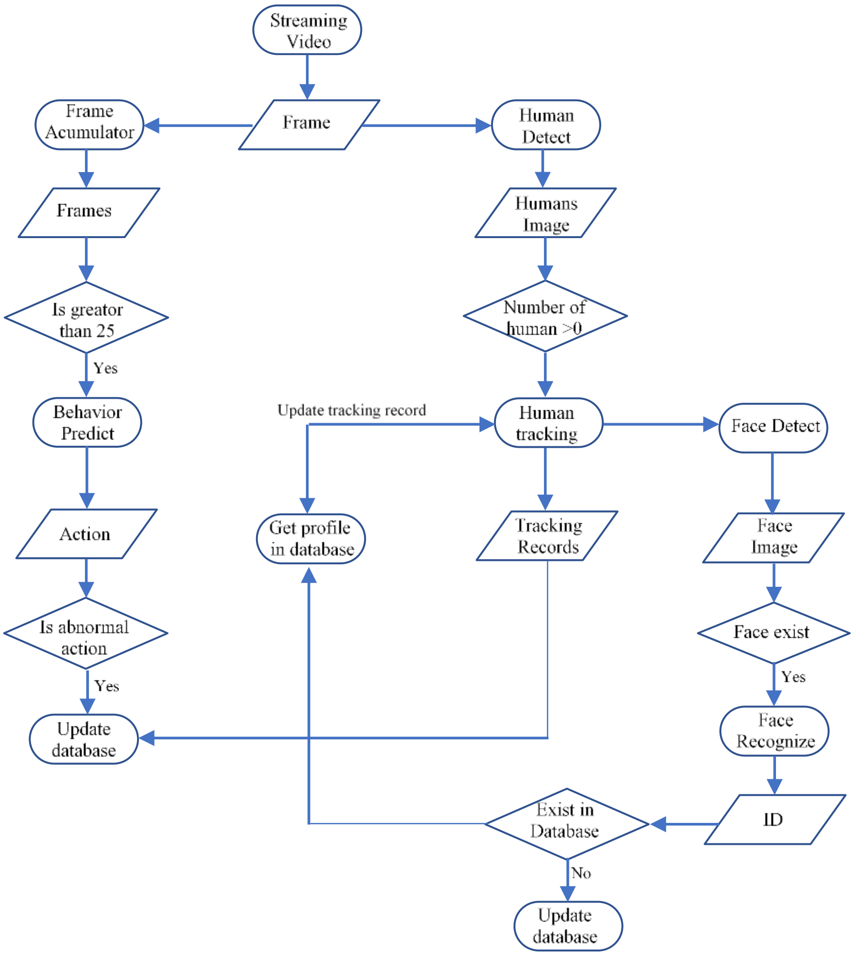

Figure 2. Follow the chart of the proposed surveillance monitoring forest model

3.2. The proposed deep learning model

The architecture of the proposed deep learning model is performed after every 15 frames. After 15 frames, a 25×2048 matrix was created from 20 features vectors of 25 frames over the InceptionV3 network. Experiments show that using 15 frames is best for balancing computational performance and behavioral classification. LSTM network structure is shown in Figure 3.

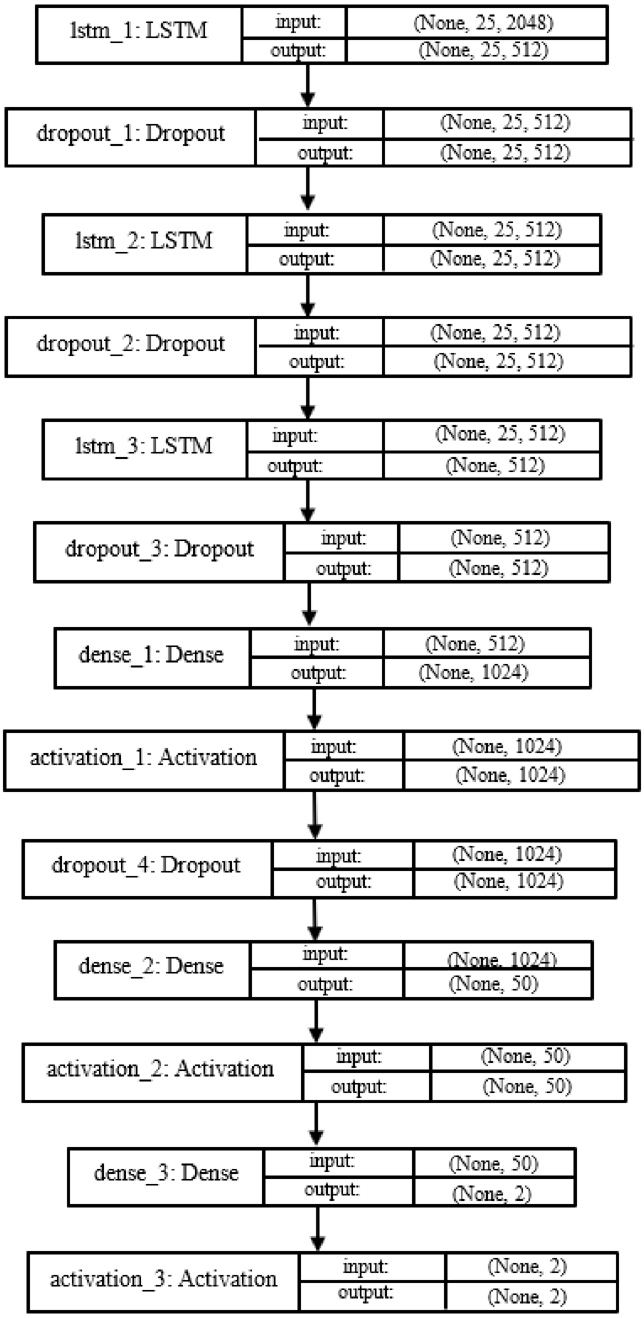

Figure 3. The architecture of the proposed deep learning model

The steps of the proposed model are described as follows. Firstly, videos from a camera have been captured in real-time. These videos are extracted into frames as images to classify images. Secondly, a series of images is transformed to LSTM network to give actions as features of human who has abnormal status in the domain of forest protection. To confirm the identification person (ID), the final step is applied to using a knowledge graph as Neo4j represented by a graph database to track human profiles.

In the system architecture, the model uses InceptionV3 shown in Figure 3 which applies to train the ImageNet series. In the experiments, both InceptionV3 and NasNetLarge models were trained on Google's ImageNet dataset. The accuracy of InceptionV3 was lower than NasNetLarge (top 1 accuracy 0.779 vs 0.825) belonging to its compact structure, InceptionV3 calculation speed was faster than NasNetLarge. The behavioral classification would be performed every 15 frames. After 15 frames, a 25×2048 matrix was created from 20 features vector of those 25 frames over the InceptionV3 network. Experiments show that using 15 frames is best for balancing computational performance and behavioral classification. LSTM network structure shows in Figure 3. Feature matrix 25×2048 was represented in the LSTM network with the output being the softmax layer, which plays the role of video classification. The experiments through several network architectures show that the model of 3 LSTM layers gives the best results. For the loss function, the model using the common loss function for the classification problem is the cross-entropy function. The output is the proposed system showed the ID person integrated with knowledge graph stored in graph database as discussed in section 3.3.

3.3. Integrated knowledge graph to deep learning model

This example has shown the sending requests from the original ID person including profile and then detecting the entity from a person with his/ her profile, who visits a forest. The main attributes of the entity are identified. The entity type $O = [Per, Pro, Dis, War, Lan, Loc, Wpl, Air, Tra\_sta, Doc, Bus\_sta, Bus\_sto]$ is collected. The name of the field is presented in Table 1. Each field is found and split into a corresponding entity.

Based on the response that is returned from the initial request uniform resource locators (URL), the ReLeaSE (REL) entities are detected and moved into set R to establish the relationships from the new entities to the original entity. The relationships between entities are described as follows. The integration of a knowledge graph is to Figure 4 out two algorithms as: 1) algorithm 1 is applied to create a knowledge graph; and 2) algorithm 2 is used to create a weighted directed relationship between person and location by calculating the total times' check-in of a person at the location of a forest. Multidimensional inference between people and places is presented in Table 2.

Table 1. Entities name notation

Symbols	Meanings
Per	Persons (name, ID, gender, identification) are generated randomly
Pro	Provinces/Cities (63 nodes)
Dis	Districts (709 nodes)
War	Wards (11162 nodes)
Lan	Landmarks are generated randomly
Loc	Locations are in the forest
Wpl	Workplaces in a forest
Air	Airports (22 nodes)
Tra_sta	Train stations (92 nodes)
Doc	Docks forest (9 nodes)
Bus_sta	Bus stations (424 nodes)
Bus_sto	Bus stops are in the forest

Table 2. Multidimensional inference between people and places

Image explanation
(Province)---[HAVE]→(District)
(Province)---[HAVE]→(Airport)
(Province)---[HAVE]→(Train station)
(Province)---[HAVE]→(Bus station)
(Province)---[HAVE]→(Dock)
(District)---[HAVE]→(Ward)
(Ward)---[HAVE]→(Landmark)
(Ward)---[HAVE]→(Location)
(Ward)←[LOCATE]---(Bus stop)
(Ward)←[LOCATE]---(Work space)
(Person)---[LIVE]→(Ward)
(Person)---[WORK]→(Work place)
(Person)---[CHECK_IN]→(Location)
(Person)---[CHECK_IN]→(Landmark)
(Person)---[CHECK_IN]→(Airport)
(Person)---[CHECK_IN]→(Train station)
(Person)---[CHECK_IN]→(Bus station)
(Person)---[CHECK_IN]→(Dock)
(Person)---[CHECK_IN]→(Bus stop)

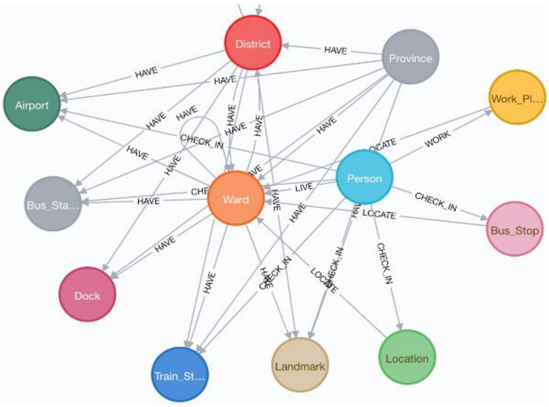

Figure 4. A knowledge graph about the relationship between the person and places

Algorithm 1. Create a knowledge graph

```
For each object in $Objects$
    If object is in [$Per$, $Pro$, $Dis$, $War$, $Lan$, $Loc$, $Wpl$, $Air$, $Tra\_sta$, $Doc$, $Bus\_sta$, $Bus\_sto$]:
        $NodeG$ = {}
        $NodeG$ += object attributes
        $NodeG$ += type of object
        Create $NodeG$ in graph
For each Rel in relationship:
    $RelG$ = {}
    For each object in Rel:
        $RelG$ += object
    $RelG$ += $R$ type attributes
    Create $RelG$ in graph
```

Algorithm 2. Create a weighted directed relationship between person and location by calculating the total time's check-in of a person at the location of a forest

Input: list the status of the location

Output: weighted relationships between person and location

```
Foreach status in $list\_people$
    Place = get a place of status
    $list\_res$ = get response of status
    Foreach res in $list\_res$
        Person = get person object of response
        $rel\_person\_place$ = get relationship interact from person to place
        If exist $rel\_person\_place$:
            Weight = $get\ weight\ of\ rel\_person\_place$
        Else:
            $is\_like\_place$ = get relationship person like place
            If exist $is\_like\_place$:
                Weight =1
            Else:
                Weight = 0
            $rel\_person\_place$ = {weight: weight}
        If $respone.type$ == 'abnormal'
            Weight += 0.75
        Else if $respone.type$! = 'abnormal' and $respone.type$! = 'warning alarm':
            Weight += 0.9
        Else:
            Weight += 1
        $rel\_person\_place$ = {weight: weight}
Model schema ($rel\_person\_place$)
```

4. EXPERIMENTAL RESULTS

4.1. Data sets, experiments, and cases study

The proposed model has been tested using a dataset of the Asian celebrity set provided by DeepGlint [26]. The dataset includes 93,979 identities out of a total of 2,830,146 processed images that identify face detection. From the above image series, the proposed model has been tested on all faces by the landmark file containing the face coordinates in the image and resized to 112×112. In the experiments, some training parameters of the proposed model are as:

- Training data: Asian celebrity includes 93,000 identities / 2.8 M photos.
- Hardware: 2 GPU Tesla P100.
- Batch size: 64.
- Optimal algorithm: gradient momentum.
- Epoch number: 14.

In a case study of forest protection in this video, the human has completely normal behavior in the proposed system has been recognized as a "normal" prediction, as shown in Figure 5. It also shows the results of the person in the graph database so we can check a person's historical profile, as shown in Figure 6. In this video, the proposed model shows a behavioral surveillance person who is detected with its statuses such as abnormal behavior (cutting down the tree) and the face recognition model detects this person's identity, so the action about this person's behavior also adds his abnormal behavior to the proposed system. Another example shows all features of a human who visits the forest. It is possible to track the person in the Knowledge graph as Neo4J as shown in Figure 7.

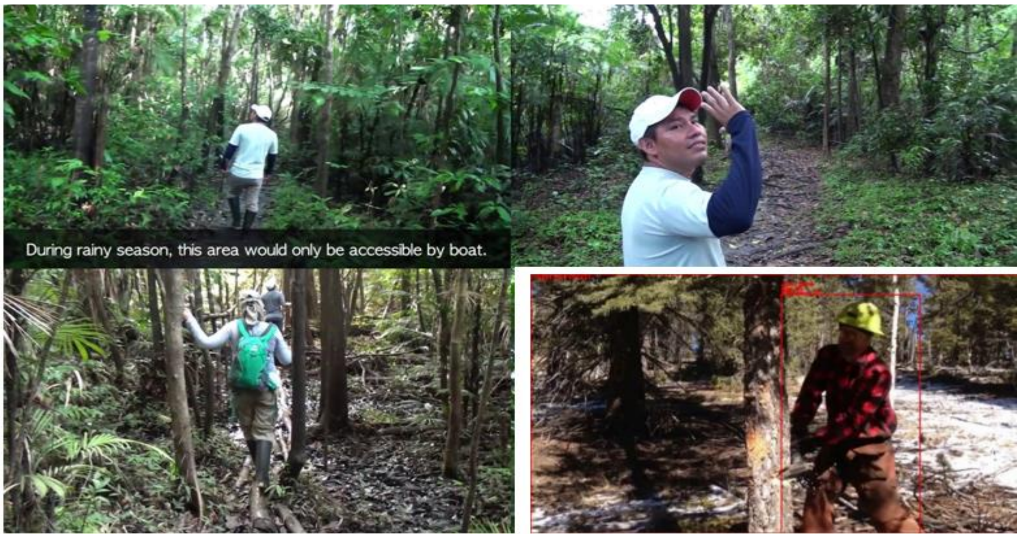
During rainy season, this area would only be accessible by boat.

Figure 5. Warning abnormal behavior cutting the tree

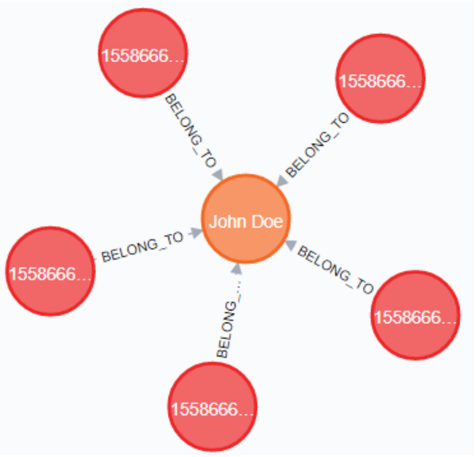

Figure 6. "John Doe" abnormal behavior profile in the graph database

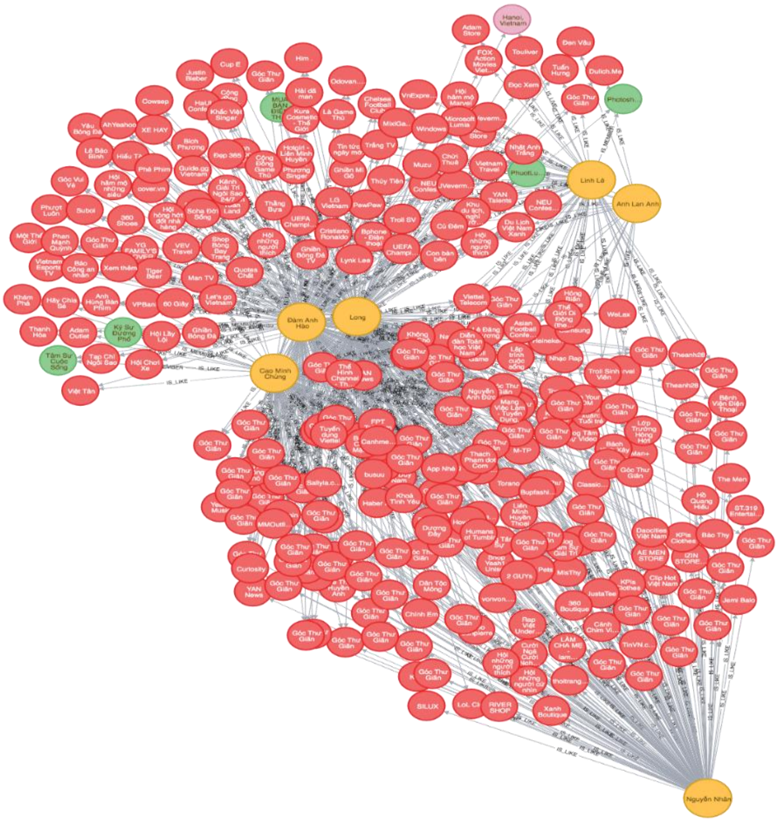

Figure 7. Abnormal behavior profile in graph database of knowledge graph

4.2. Result Discussions

To validate the proposed model, the proposed model has been tested with two methods such as benchmark of large-scale unconstrained face recognition method (BLUFR) and behavior monitoring (entropy cross) using cameras taking a series of images in real-time in the forest.

1) BLUFR evaluation: the evaluation method is used to evaluate pairs of images. The proposed model has been tested through a case study to make the right decision of human recognition which pair of images of persons. In the experiments, the parameters set, as expressed by (1), (2) in the following parameters.

$$TA(d) = \{(i,j) \in P_{same}, với\ D(x_i, x_j) \leq d\} \quad (1)$$

$$FA(d) = \{(i,j) \in P_{diff}, với\ D(x_i, x_j) \leq d\} \quad (2)$$

Where TA represents a true value which is distance pairs of the same as identity measurement. FA represents a false value that represents distance pairs of various identities when misclassified. d is the threshold that determines vectors whether belong to the same identity or not. To validate the rate, it is expressed by (3), (4).

$$VAL(d) = \frac{|TA(d)|}{P_{same}} \quad (3)$$

$$FAR(d) = \frac{|FA(d)|}{P_{diff}} \quad (4)$$

The accuracy is the number of images that are correctly verified over the total number of images. The prosed model has been tested by using the BLUFR. The model has been tested by the BLUFR method with minimizing the cases of false recognition, setting the false acceptance ratio (FAR) value at 0.1%. To evaluate the training results of the proposed model, it was compared with another pre-training model:

– Training data: MS1M-ArcFace includes 85,000 identities / 5.8M images (data source [14], [15]).
– Hardware: 8 Tesla P40 GPU. Batch Size: 256.

The proposed model in evaluation results by using the BLUFR method using datasets as:

– Labeled faces in the Wild (LFW): 5,749 identities / 13,233 images / 6000 pairs (source: item [16] references).
– AgeDB-30: 570 identities / 12,240 images / 6000 pairs. It is a diverse dataset of age (source: item [17] references).
– Celebrities in frontal and profile (CFP): 500 identities / 7,000 and self-collection: 67 identities / 735 images / 6360 pairs consist of frontal and cross-sectional photos.
– The celebrity dataset was taken in 2 types horizontal and frontal (source: item [18] references).
  a) CFP-FP: 7000 pairs. Including 1 frontal photo and one horizontal photo.
  b) CFP-FF: 7000 pairs. Includes pairs of frontal photos.

2) Behavior monitoring evaluation

The evaluation method of the proposed model has been calculated by the percentages of accuracy by dividing the total number of correct predictions by the amount of data. The model loss function used is a “categorical cross-entropy” with the purpose mostly based on the number of data layers. While calculating the value of the loss function, the labels of data for the model training are encoded as “one-hot encoding”.

$$y_{onehot} = [y_1, y_2, \dots, y_N] \begin{cases} y_i = 1 \\ y_j = 0 \ \forall j \neq i \end{cases} \quad (5)$$

Where i is the actual label of data, and N is the number of data layers.

The output of the softmax function is a vector with N dimensions representing the probability of data point that belongs to different layers.

$$y_{predict} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_N] \quad (6)$$

The purpose of the categorical cross-entropy is to penalize other predictions that are different from the actual label of the data point. Thus, the “entropy” function gives a percentage of similar two vectors of probability distribution in the same direction as expressed by (7).

$$entropy = -\sum_{i=0}^{N} y_i \log(x_i), x_i \in X, y_i \in Y \quad (7)$$

The entropy value would be large when the two probabilities are considered. As shown in the formula for the categorical cross-entropy loss function by summing the entropy values of the given predictions with its one-hot vector as shown in (8).

$$L(y, \hat{y}) = -\sum_{j=0}^{M} \sum_{i=0}^{N} y_{ij} \log(\hat{y}_{ij}) \quad (8)$$

Where M is the number of data points and N is the number of data layers. For the “batch gradient” method, M is the number of data points in the batch, which is equivalent to the batch-size value. Experiments of the proposed model have been tested using 5524 data points. The test has been divided by 8:2 into the data set for training and testing. The LSTM model was trained based on a cross-validation strategy. Training data is shuffled and divided in a 9:1 ratio for training and validation. As shown in Table 3, a comparison of training results for the CNN + LSTM model indicates. As shown in Table 3, the proposed model (NasNetLarge + 3 × LSTM (512) + FC (1024) + FC (50)) in experimental results indicate that the proposed model has been validated on real-world datasets to demonstrate this method’s effectiveness.

Table 3. Comparison of training results for CNN + LSTM model

Architect	Train loss	Train accuracy (%)	Validation loss	Validation accuracy (%)	Test loss	Test accuracy (%)
InceptionV3 + 3 × LSTM (512) + FC (1024) + FC (50)	0.0945	96.66	0.1263	95.55	0.2006	93.49
InceptionV3 + 2 × LSTM (512) + FC (1024) + FC (50)	0.1843	93.6	0.1286	95.18	0.1935	93.15
InceptionV3 + 3 × LSTM (512) + FC (768) + FC (50)	0.1812	93.72	0.1563	94.98	0.1744	93.6
NasNetLarge + 3 × LSTM (512) + FC (1024) + FC (50)	0.0644	97.43	0.0847	96.79	0.1047	96.38

5. CONCLUSION

In this paper, we have presented a new method for the improvement proposed deep learning model integrated with a knowledge graph with an adaptive prioritization mechanism for the surveillance monitoring system to track human behavior in real-time for the forest protection domain. To address a range of typical situations we use dynamic questions and responses based on discussions with advice from experts and consultants. Experimental results indicate that the theoretical basis of deep learning integrated with a graph database to demonstrate human behaviors by tracking human profiles to apply forest protection using this method's effectiveness. The proposed model proves a novel approach using deep learning for face recognition with its behavioral surveillance of the human profile integrated with a graph database that can be applied in real-time in a forest protection domain. For further investigation in this study, it should be extended the models of Deep learning integrated with knowledge graphs in reasoning to track groups of human behaviors and relational activities of human groups in real-time.

ACKNOWLEDGEMENTS

This research is funded by University of Economics Ho Chi Minh City (UEH), Vietnam.

REFERENCES

[1] C. Yuan, Z. Liu, and Y. Zhang, "Vision-based forest fire detection in aerial images for firefighting using UAVs," in *2016 International Conference on Unmanned Aircraft Systems (ICUAS)*, 2016, pp. 1200-1205, doi: 10.1109/ICUAS.2016.7502546.

[2] C. Yuan, Z. Liu and Y. Zhang, "UAV-based forest fire detection and tracking using image processing techniques," *2015 International Conference on Unmanned Aircraft Systems (ICUAS)*, 2015, pp. 639-643, doi: 10.1109/ICUAS.2015.7152345.

[3] S. Sudhakar, V. Vijayakumar, C. S. Kumar, V. Priya, L. Ravi, and V. Subramaniyaswamy, "Unmanned Aerial Vehicle (UAV) based Forest Fire Detection and monitoring for reducing false alarms in forest-fires," *Computer Communications*, vol. 149, pp. 1-16, 2020, doi: 10.1016/j.comcom.2019.10.007.

[4] F. Sharifi, Y. Zhang, and A. G. Aghdam, "Forest fire monitoring and detection using a network of autonomous vehicles," in *Proceedings of International Conference on Intelligent Unmanned Systems*, 2014, vol. 10, pp. 190-210. [Online]. Available: http://ojs.unsysdigital.com/index.php/icius/article/view/302

[5] P. Moore and H. Van Pham, "On Context and the Open World Assumption," *2015 IEEE 29th International Conference on Advanced Information Networking and Applications Workshops*, 2015, pp. 387-392, doi: 10.1109/WAINA.2015.7.

[6] R. Poppe, "A survey on vision-based human action recognition," *Image vision computing Journal*, vol. 28, no. 6, pp. 976-990, 2010, doi: 10.1016/j.imavis.2009.11.014.

[7] S. Han and S. Lee, "A vision-based motion capture and recognition framework for behavior-based safety management," *Automation in Construction,* vol. 35, pp. 131-141, 2013, doi: 10.1016/j.autcon.2013.05.001.

[8] N. Zerrouki, F. Harrou, Y. Sun and A. Houacine, "Vision-Based Human Action Classification Using Adaptive Boosting Algorithm," in *IEEE Sensors Journal*, vol. 18, no. 12, pp. 5115-5121, 2018, doi: 10.1109/JSEN.2018.2830743.

[9] L. Meng, "Design of Forest Fire Detection Algorithm Based on Machine Vision," *2021 International Conference on Electronic Information Technology and Smart Agriculture (ICEITSA)*, 2021, pp. 117-121, doi: 10.1109/ICEITSA54226.2021.00031.

[10] P. Moore and H. V. Pham, "Intelligent Context with Decision Support under Uncertainty," *2012 Sixth International Conference on Complex, Intelligent, and Software Intensive Systems*, 2012, pp. 977-982, doi: 10.1109/CISIS.2012.17.

[11] H. V. Pham and V. T. Nguyen, "A novel approach using context matching algorithm and knowledge inference for user identification in social networks," in *Proceedings of the 4th International Conference on Machine Learning and Soft Computing*, 2020, pp. 149-153, doi: 10.1145/3380688.3380708.

[12] H. V. Pham, P. Moore, and K. D. Tran, "Context matching with reasoning and decision support using hedge algebra with Kansei evaluation," in *Proceedings of the Fifth Symposium on Information and Communication Technology*, 2014, pp. 202-210, doi: 10.1145/2676585.2676598.

[13] V. B. S. Prasath, D. N. H. Thanh, N. Q. Hung and L. M. Hieu, "Multiscale Gradient Maps Augmented Fisher Information-Based Image Edge Detection," in *IEEE Access*, vol. 8, pp. 141104-141110, 2020, doi: 10.1109/ACCESS.2020.3013888.

[14] H. V. Pham and Q. H. Nguyen, "Intelligent IoT Monitoring System Using Rule-Based for Decision Supports in Fired Forest Images," in *International Conference on Industrial Networks and Intelligent Systems*, 2021, vol. 379, pp. 367-378, doi: 10.1007/978-3-030-77424-0_30.

[15] L. T. Thanh, D. N. H. Thanh, N. M. Hue and V. B. S. Prasath, "Single Image Dehazing Based on Adaptive Histogram Equalization and Linearization of Gamma Correction," *2019 25th Asia-Pacific Conference on Communications (APCC)*, 2019, pp. 36-40, doi: 10.1109/APCC47188.2019.9026457.

[16] L. T. Thanh and D. N. H. Thanh, "An adaptive local thresholding roads segmentation method for satellite aerial images with normalized HSV and lab color models," in *Intelligent Computing in Engineering*, 2020, vol. 1125, pp. 865-872, doi: 10.1007/978-981-15-2780-7_92.

[17] M. Norris, R. Anderson, and I. C. Kenny, "Method analysis of accelerometers and gyroscopes in running gait: A systematic review," *Sage journals*, vol. 228, no. 1, pp. 3-15, 2014, doi: 10.1177/1754337113502472.

[18] L. Fan, Z. Wang and H. Wang, "Human Activity Recognition Model Based on Decision Tree," *2013 International Conference on Advanced Cloud and Big Data*, 2013, pp. 64-68, doi: 10.1109/CBD.2013.19.

[19] A. Filippeschi, N. Schmitz, M. Miezal, G. Bleser, E. Ruffaldi, and D. Stricker, "Survey of motion tracking methods based on inertial sensors: A focus on upper limb human motion," *Sensors*, vol. 17, no. 6, 2017, doi: 10.3390/s17061257.

[20] B. Su, D. Zheng, and M. Sheng, "Single-sensor daily behavior recognition based on functional data time series modeling," *Pattern Recognition and Artificial Intelligence,* vol. 31, pp. 653-661, 2018, doi: 10.16451/j.cnki.issn1003-6059.201807008.

[21] A. Graves, "Long short-term memory," in *Supervised sequence labelling with recurrent neural networks*, Berlin, Heidelberg: Springer, 2012, vol. 385, pp. 37-45, doi: 10.1007/978-3-642-24797-2_4.

[22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Journal of Neural computation,* vol. 9, no. 8, pp. 1735-1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[23] Z. Yu and W. Q. Yan, "Human Action Recognition Using Deep Learning Methods," *2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ)*, 2020, pp. 1-6, doi: 10.1109/IVCNZ51579.2020.9290594.
[24] N. Tufek, M. Yalcin, M. Altintas, F. Kalaoglu, Y. Li and S. K. Bahadir, "Human Action Recognition Using Deep Learning Methods on Limited Sensory Data," in *IEEE Sensors Journal*, vol. 20, no. 6, pp. 3101-3112, 2020, doi: 10.1109/JSEN.2019.2956901.
[25] J. Lu, M. Nguyen, and W. Q. Yan, "Deep Learning Methods for Human Behavior Recognition," *2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ)*, 2020, pp. 1-6, doi: 10.1109/IVCNZ51579.2020.9290640.
[26] *Trillion Pairs Dataset*, Deepglint, 2018. [Online]. Available: http://trillionpairs.deepglint.com/overview

BIOGRAPHIES OF AUTHORS

Van Hai Pham is received Doctor of Engineering degree (Ph.D.) at Ritsumeikan University (Japan). He is an Associate Professor at School of Information and Communication Technology, Hanoi University of Science and Technology. His major fields include Artificial Intelligence, Knowledge Based, Big data, Soft Computing, Rule-based Systems and Fuzzy Systems. His an Associate Editor of International Journal of Computer Applications in Technology and other domestic and International journals. He also serves as Chairs and Co-chairs of organized several sessions at international conferences such as KSE 2019, KSE 2017, KSE 2015, SOICT 2014. He can be contacted at email: haipv@soict.hust.edu.vn.

Quoc Hung Nguyen is the Senior Lecturer/Researcher at the School of Business Information Technology (BIT), University of Economics Ho Chi Minh City (UEH). He received Ph.D. in Computer Science of Hanoi University of Science and Technology (HUST), in 2016. His research interests include methods for big data analytics using artificial intelligence and blockchain encryption technology in applications in the fields of economics, business, science, technology, medicine, and agriculture. He has published over 07 journal papers, 5 authored books/ book chapter, and 50 papers in conference proceedings and some application results in robotics and image processing. He can be contacted at email: hungngq@ueh.edu.vn.

Thanh Trung Le is the Vice Dean of Faculty Information Technology, University of Economics Ho Chi Minh City (UEH), Vietnam. He received a Master’s degree in Information Systems of CanTho University, Faculty of Information and Communication Technology (ICT), in 2011. His research interests include Data mining, Machine Learning, artificial intelligence, and blockchain. He can be contacted at email: trunglt@ueh.edu.vn.

Thi Xuan Dao Nguyen is the Deputy Head of Faculty Information Technology, University of Economics Ho Chi Minh City (UEH), Vietnam. She received a master’s degree in Information Systems from Can Tho University. Her research interests include data analysis, Database Systems, and Data Mining. She can be contacted at email: daontx@ueh.edu.vn.

Thi Thuy Kieu Phan is a lecturer/researcher of Faculty Information Technology, University of Economics Ho Chi Minh City (UEH), Vietnam. She received a master’s degree in Information Systems from Can Tho University. Her research interests include Machine Learning, Data Mining, Information Extraction, Information Retrieval, and AI. She can be contacted at email: kieuptt@ueh.edu.vn.