A computer vision framework for detecting and preventing Human-Elephant Collisions

Pushkar Shukla
University of California , Santa Barbara
pushkarshukla@umail.ucsb.org

Isha Dua
International Institute of Technology, Hyderabad
isha.dua@research.iiit.ac.in

Balasubramanian Raman
Indian Institute of Technology, Roorkee
balarfma@iitr.ac.in

Ankush Mittal
Graphic Era University, Dehradun
dr.ankush.mittal@gmail.com

Abstract

Human Elephant Collision (HEC) is a problem that is quite common across many parts of the world. There have been many incidents in the past where conflict between humans and elephants has caused serious damage and resulted in the loss of lives as well as property. The paper proposes a frame-work that relies on computer vision approaches for detecting and preventing HEC. The technique initially recognizes the areas of conflict where accidents are most likely to occur. This is followed by elephant detection system that identifies an elephant in the video frame. Two different algorithms to detect the presence of elephants having accuracies of 98.621% and 98.667% have been proposed in the paper. The position of the elephant once detected is tracked with respect to the area of conflict with a particle filter. A warning message is displayed as soon as the position of the elephant overlaps with the area of conflict. The results of the techniques that were applied on videos were discussed in the paper

1. Introduction

Elephants are the largest mammal on land belonging to *Elephantadei* family and are generally found in different parts of Africa and South East Asia. Different species of the elephant can be broadly classified on the basis of their origin as *Loxodota* (African) and *Elephas* (Asiatic) elephants. They are herbivores that prefer to stay near water bodies. Female elephant and children tend to move in groups whereas males leave the group as soon as they reach puberty. Areas having high population of elephants in South East Asia have been shown in Figure 1.

Human elephant collision is a growing area of concern especially in many parts of South East Asia. Elephants re-

Figure 1. High elephant density regions in South East Asia[1]

quire a greater amount of food and a larger ecosystem to thrive and survive. The increase in the area of human settlements has resulted into the shrinkage of forests and wild lands that serve as home for the elephant. As a result the elephant population has gone down significantly in the past few years. The loss of habitat for elephant has resulted into common zones where humans and elephants coexist. The shortage of food for elephants has led to crop raiding in the nearby agricultural fields. This not only causes financial loss to the farmers of the fields but also endangers the lives of the worker working in the field. The lives of elephants are also at risks whenever there is a conflict between elephants and humans. Each year many elephants die on railway tracks after being hit by a train. Human-elephant conflicts on roads also result in fatalities on both the sides. Therefore, taking into consideration the aforementioned problems caused due to collision of humans and elephants the development of a system that prevents conflict between humans and elephants has become a necessity in many parts of the world. Developing a good Human-

Elephant Collision (HEC) detection system has its own set of challenges. One of the biggest challenge while developing such as system would be the systems capability to recognize elephant with high accuracies. Such a system must be extremely robust to noise and the detected elephant must not be confused with vehicles, objects, humans or other animals etc. Moreover, the system should be able to identify the areas that are more prone to human-elephant collision. Further, these systems should also have the capacity to predict the collision before hand in order to prevent the collision. As soon as a potential threat is detected, an adequate warning message must be sent to the concerned authorities.

The paper proposes a computer-vision frame-work for preventing human-elephant collision. The complete framework can be broken down into the following parts.

- Identifying the area of conflict.
- Recognizing elephants in the video frame.
- Checking if the elephant is near the area of conflict.

The proposed approach relies on computer vision approaches and can be extended to real time scenarios.The major advantage of using vision based approaches over other sensory approaches is that camera based systems are more reliable when it comes to identfying and tracking elephants than sensory approaches.

The paper has been divided into five sections. Section 1. gives a basic introduction about the need of a Human- Elephant detection system. Section 2 gives a detailed survey of the existing approaches that have been used for detecting intrusion of elephant and other animals. The overall architecture of the system has been discussed in great detail in Section 3. All the experimental results have been presented in Section 4. Section 5 concludes the paper and discusses the future scope of the proposed technique.

2. Related Work

The conflict between humans and animals has increased drastically in the past few years due to the large scale destruction of animal habitat. As a result there has been an increase in the total causalities on both sides. Thereafter, many techniques have been proposed to counter this problem. Animal Detection techniques can broadly be classified into two types. Visual approaches that make use of camera or other vision sensor such as Kinect to detect animals and sensory approaches that employ some other sensor for detecting animal intrusion.

An algorithm making using Haar like features along with Adaboost classifier was designed to detect faces of animals in videos in order to recognize their locomotive behavior by [4]. [17] have used condensation particle filtering algorithm to track down the movements of multiple animals. An animal gait identification model was proposed by Hanumma et al[9]. Mammeri et al .[13] proposed a two -step classification system based on LBP-Adaboost classifier followed by a HOG-SVM classifier in order to detect the presence of moose on roads. A temporal model for animal detection that takes into account temporal coherency has been proposed by Ramanan and Forsyth[14]. A robust animal tracking system was proposed by [7]. This system used a combination of four features to identify an animal followed by tracking its location in the video.

Several techniques have been proposed to detect the presence of elephants in areas having high human intervention. An Elephant Intrusion detection System (EIDS) using Haar Features was proposed by [16] to detect and prevent collision between humans and Elephants. A message warning the official is sent, as soon as the Elephant is detected by the system. Debrera and Rodrigo [5] proposed a framework which was capable of detecting elephants and ran on a web server. [18] proposed a method to detect the position of African Elephants using satellite images. Elephant being the largest mammal are easily visible on satellite images. Therefore the elephant once detected, was tracked with a GPS sensor. The method is inefficient in areas having dense forest covers as it would be difficult to locate an elephant. Attaching an extra sensor to the animal may also make it more irritable with a chance that the senor might be damaged. Vermulen et al [19] tracked elephants with the help of images captured from unmanned aerial vehicles at a height of 100 ms above the ground level. Many detection strategies use different body parts of an Elephant to identify the animal.

[8] have used techniques to record changes in the images lobe, tusk and ear fold of different elephants. Ardovini et. al. [3] made use of images of elephant ears in order to build an identification system to classify elephants. Beehives have been proposed by [11] to counter the movements of elephants in farmlands in order to prevent crop degradation. The method might be well suited for crop fields but it cannot be applied to other areas having high elephant intervention due to issues of scalability and also the restrictions it imposes on humans. The most similar work to our approach has been proposed by [6] . They have proposed a Human-elephant collision detection system that used PHOG-SVM classifier for elephant recognition followed by a particle filter to track the elephant.

3. Methodology

This section gives a detailed description of the technique proposed to counter the problem of HEC. The overall block diagram representation has been given in Figure 2. A part of the architecture focuses on identifying regions in a video frame that are more prone to HEC . These areas of conflict are identified beforehand by localizing the coordinates of the area of conflict from the video frame followed by seg-

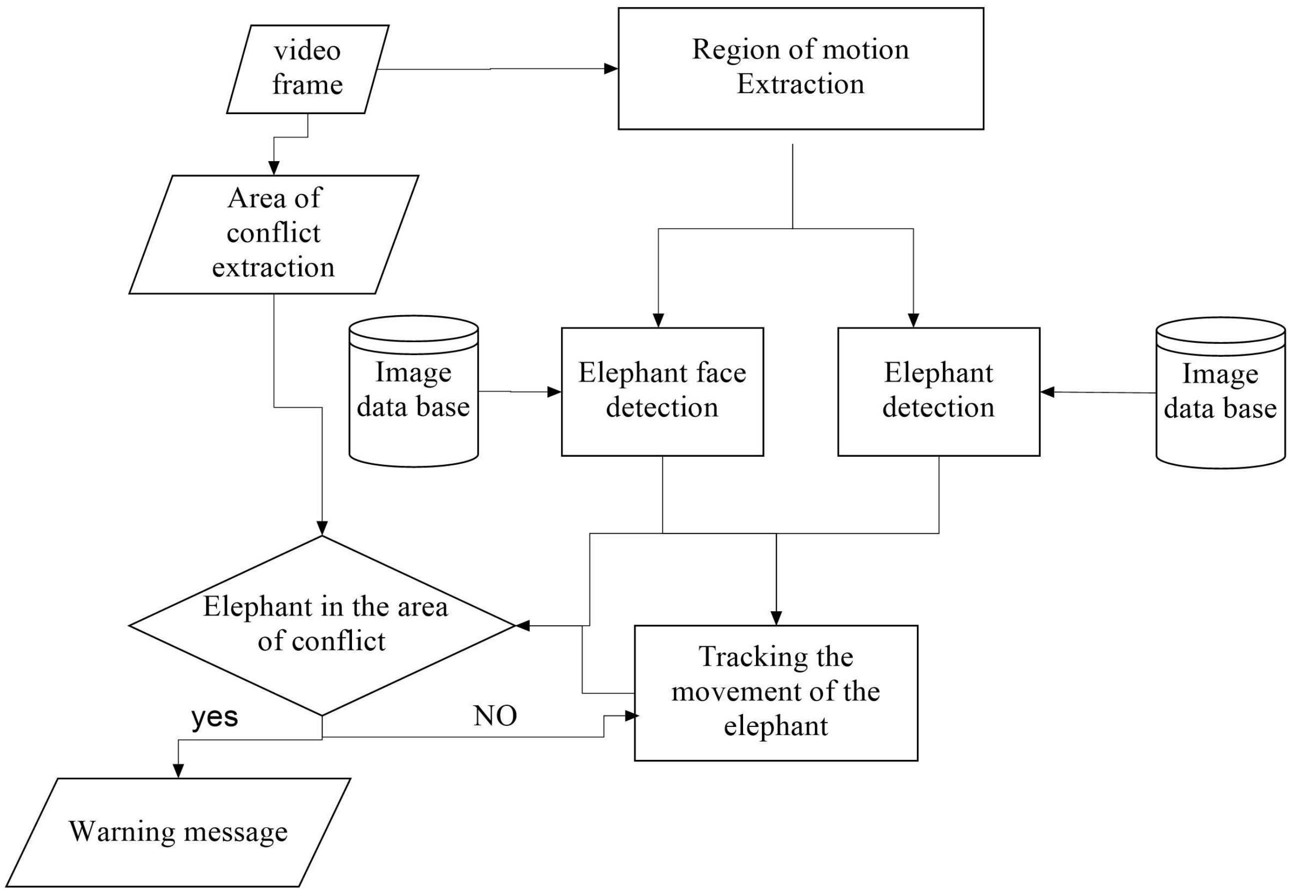

Figure 2. The diagrammatic representation of the proposed approach

mentation using the superpixel [2] technique. The other part focuses on recognizing and tracking down the positions of elephants in a video frame. An elephant must be recognized in a video frame in order to track down its movement. Therefore, the prime regions of motion are extracted from all the frames. These areas of motion are then passed onto to frameworks relying on Deep convolutional neural networks to detect the presence of elephants in those areas. Two separate frameworks are designed for elephant recognition. The first framework is trained to recognize elephants. Whereas the second framework was trained for recognizing elephant faces. The major intuition behind using two separate frameworks is that, an elephant may occupy a large part of the video frame, therefore, the segmentation algorithm might not be able to capture the complete elephant.

If the elephant is detected in a given area, then its position is tracked with respect to the area of conflict. A warning message is generated if the elephant is in the area of conflict. Otherwise, the movement of the elephant is continuously monitored.

3.1. Extracting the area of Conflict.

The area of conflict can be defined as the region in the video frame that has the highest probability of collisions between humans and elephants. The areas of conflict will therefore differ from region to region and from video to video. It has been assumed that in the proposed system the position of the camera is fixed. Therefore, it may capture regions having a low probability of humans colliding with elephants as well as regions having a high probability of human elephant collisions. The whole scene captured by the video camera can thus be divided in to two parts, depending on whether the human elephant collision is likely to happen in that area or not. Hence, there is a need to identify areas having higher likelihood of human-elephant collision. These areas generally include roads or railway tracks. Agricultural lands are also an area of conflict as they are often raided by elephants for food.

A superpixel[2] based segmentation approach was applied for extracting the region of conflict. The input frame was segmented into H clusters using the superpixel algorithm. Another approach that relied on color and spatial

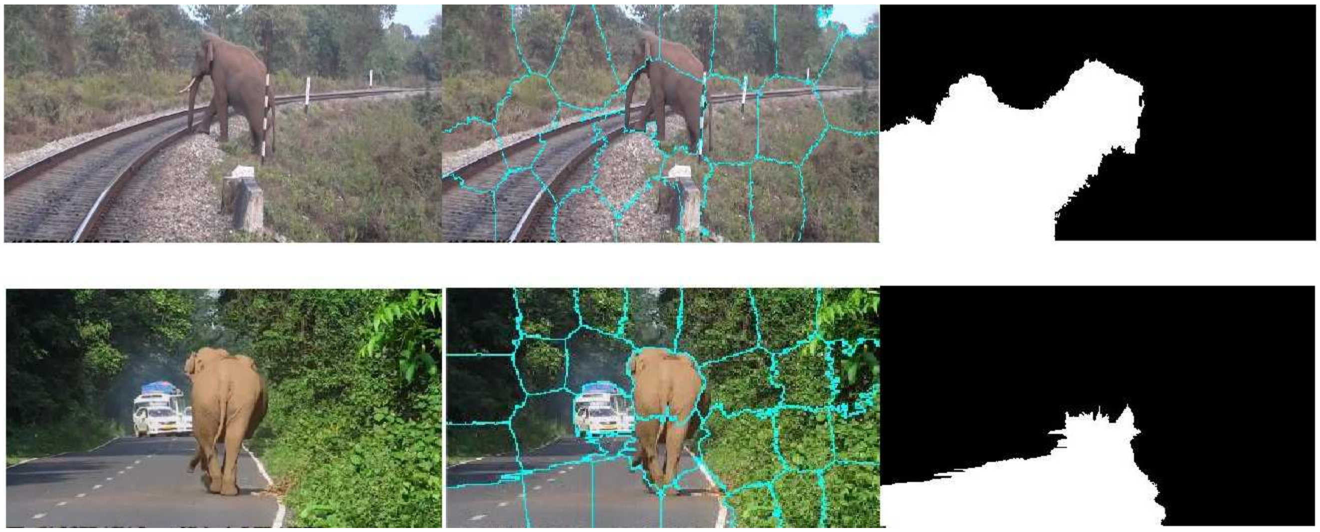

Figure 3. Extracting the area of conflict for two different scenarios. The railway track was identified as the area of conflict in the first scenario while the road was identified as the area of conflict in the second scenario. The coordinates of the area of conflicts that were extracted using the proposed approach have been shown in the figure along with super pixel cluster for the frame.

properties of the region was also used for extracting the region of conflict.The segmented super-pixel cluster that overlapped sufficiently with region of conflict extracted earlier by using the color and spatial properties were now labeled as potential regions of conflict, if the overlapping ratio was greater than 15% of the total number of pixels in the cluster. The process of extracting the regions of conflict has been demonstrated in Figure 3.

3.2. Identifying the major areas of motions

The elephant must be detected in a video frame to check for the possibilities of HEC. The position of the elephant can be tracked once it has been detected. We have assumed that the elephant will be in a state of motion at some point of time in the video. One approach to check for the presence of an elephant in a video frame is to detect the areas of motion in the video and then check whether the elephant is present in those areas of motion or not. Therefore, a technique was proposed to identify and crop the major areas of motion in the video.

The video is sampled after every n frames. The difference image D(:,:,t) is obtained by subtracting the current frame F(:,:,t) from the previous frame F(:, :, t-n) to obtain the areas of motion

$$D(:,:,t) = F(:,:,t) - F(:,:,t-n) \tag{1}$$

The obtained image is then converted into its binary equivalent followed by morphological operations to remove the noise present in the binary image. Then the binary image is further divided into *y* different cells. The value of *y* in our system was taken to be 4. The motion of the image is now captured with the help of a matrix *y*y* dimensional matrix *Z*, such that each element of *Z* is the sum of all the elements of the corresponding cell *C* in the binary image as shown in Eq.2.

$$Z(k,l) = \sum C(k,l) \tag{2}$$

A Binary Pattern for Motion (BPM) is then generated by thresholding Z where the values of all cells are set to 1 if they are greater than the threshold and they are set 0 if they are less than the threshold. The BPM for a single image has been shown in Figure 4.

$$Z(k,l) = \sigma C(k,l) \tag{3}$$

The value of threshold was chosen to be 50 in the proposed system. Figure 4 shows the binary pattern for motion obtained for a particular difference image. The object of interest may lie in either of these cells or a combination of these cells. So all the sequences in the binary pattern for motion having 1s adjacent to each other need to be extracted. Therefore, breath first search was used for this purpose.

A crop matrix containing the location of different areas of motion that are adjacent to each other is obtained after applying BFS algorithm to the binary matrix for motion. The crop matrix is an n*4 matrix. Each vector of the crop matrix is 4 dimensional, containing information about the starting point and the end points of the cells enclosing the areas of motion. The crop matrix contains a lot of sequences that overlap each other. These overlapping sequences are then combined to form a fewer set of sequences. Then, areas

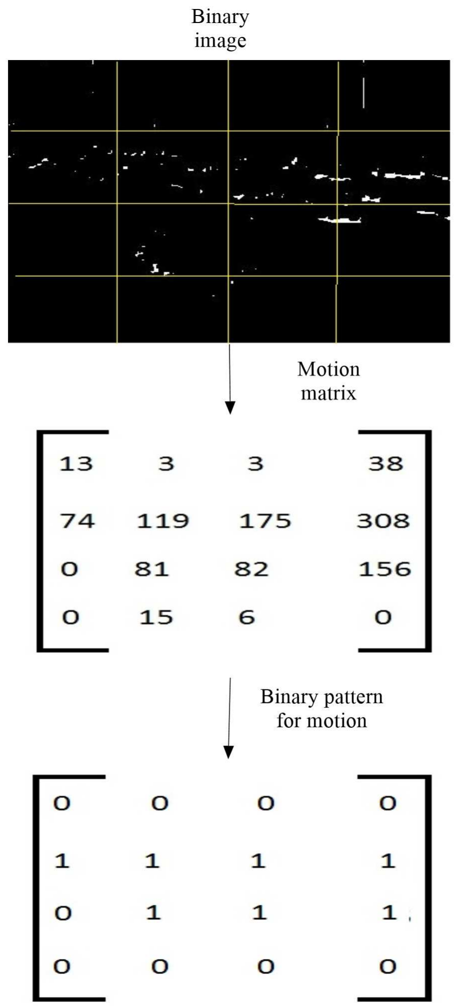

Figure 4. Generating a binary pattern for motion for a binary image containing the difference of two frames. The image was divided into 16 cells. The total number of 1's in each cell are then calculated and stored in a motion matrix. The thresholding of the motion matrix results into a binary pattern for motion.

of motion in the image are extracted from the crop matrix according to their position.

3.3. Elephant Recognition using Deep representations

The areas of motion once extracted, are now fed into an elephant detection system for recognizing whether the area of motion contains an elephant or not? Two different classification models were trained in order to have a higher recognition rate. The first model was designed to recognize the faces of an elephant in an image and the second model was designed for identifying elephants or groups of elephants in the cropped image. This was done in order to increase the efficiency of the system.The recent application of deep Neural networks in image recognition tasks was the primary reason for choosing the model[15]. The elephant was said to be detected, if any of the two models were able to recognize the elephant. Deep representations extracted from the pre-trained Alex-Net [12] model were used as descriptors. Representations were extracted from the *fc7* layer. The dimensions of a single representation was 4096*1. The success of Deep Convolutional Neural Networks for image and video recognition tasks [10, 20] was the reason for choosing them over other hand crafted descriptors.

The representations were then fed into a binary Support Vector Machine(SVM) for classification. The SVM was trained on a linear kernel with a kernel offset equal to 0. The SVM was optimized using a Sequential Minimal optimization strategy.

3.4. Tracking and Intrusion detection

The position of the elephant is tracked as soon as the elephant is detected with the aid of a particle filter. The particles associated with the elephant are a representative of the elephants position. An elephant is said to be in the area of conflict, if a large number of particles associated with the elephant overlap with the area of conflict. Based on the hypothesis,the proposed model assumes an elephant to be in the area of conflict if more than one third of the total particles that are tracking the elephant are in the region of conflict. Once the elephant was detected in the area of conflict, the color of the particles associated with the elephants changed in order to signify a warning message. As the experiments were done on prerecorded videos and not on a real time systems, an adequate warning system could not be developed.

4. Results

4.1. Data-set description

The proposed methodology was tested on a total of 12 videos. Roads ,railway tracks and agricultural lands are generally the area where elephant intrusion is very common. There are two videos of elephants crossing railway tracks, two videos where elephants are destroying agricultural crops and the remaining videos depict human elephant collision on roads. These videos were collected from youtube. It must be noted that the authors were also wanted to test the proposed model on night vision videos but unfortunately we could not find one.

The data-set for elephant face recognition comprised of images of 500 faces of elephant. Samples of images present in the data-set have been shown in Figure 6. The negative class comprises of objects generally seen on roads like vehicles, humans, and on railway tracks like trains.

The data-set for elephant face recognition comprises of 500 images of elephants in groups, on roads or in natural

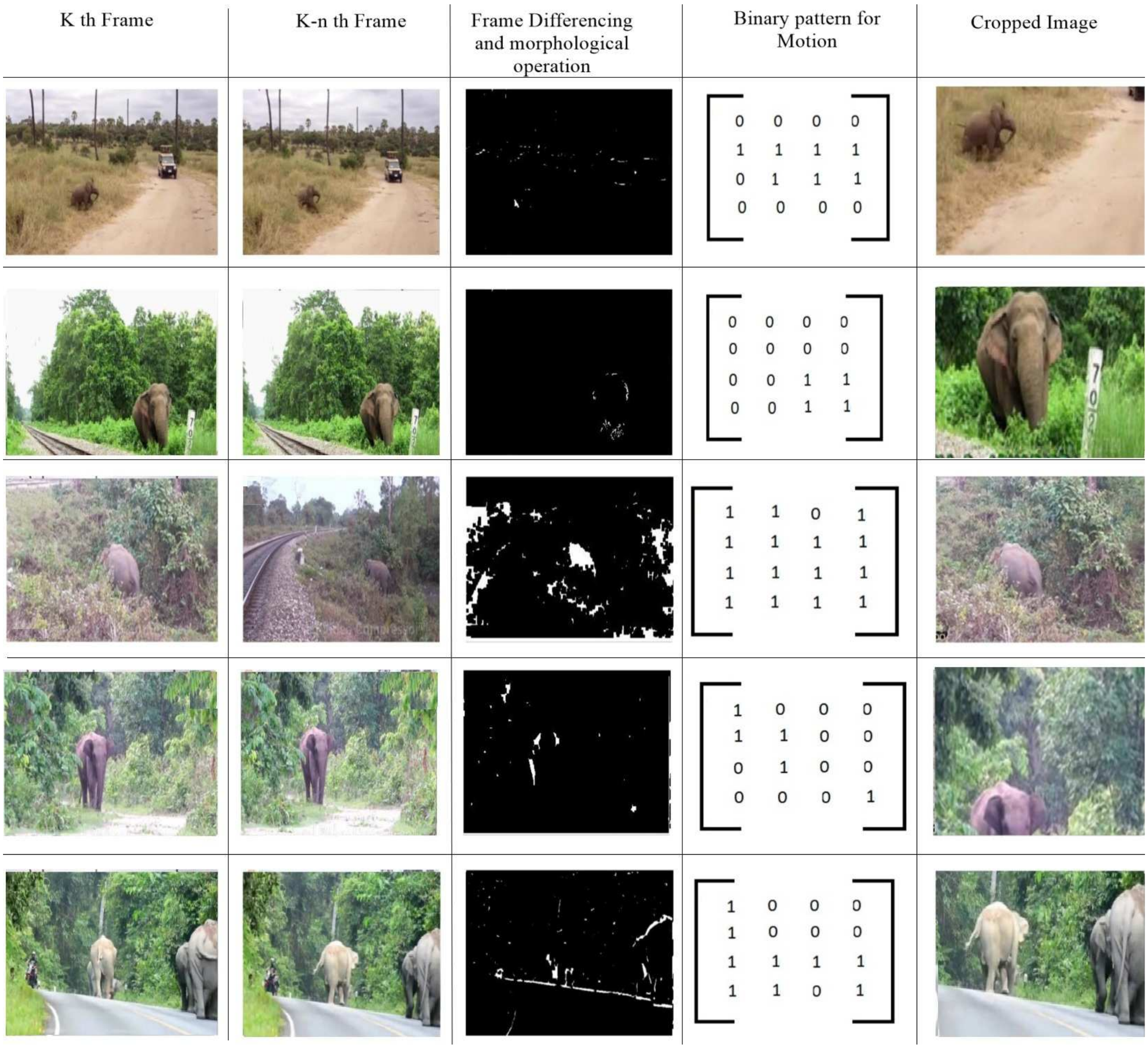

Figure 5. The different parts of the process of extracting the area of motion from different video frames in different videos. The binary pattern for motion is generated for the binary image containing the difference of the images. Later the cropped image is generated according to the binary pattern for motion.

landscapes. Samples of images present in the data-set have been shown in Figure 7. The negative class for classification is same as used for elephant face classification.

4.2. Area of motion Extraction

The area of motion containing the elephant was successfully extracted from the videos as proposed by the algorithm. Figure 5 demonstrates the aspects of the process of extracting the area of motion from different videos. The extracted image was then fed to the elephant recognition system in order to check whether the cropped image contains an elephant or not.

4.3. Elephant recognition

In this section, the results of elephant recognition algorithm was compared with several other hand-crafted and previously applied techniques for elephant recognition. The framework achieved a Mean Average Precision of 98.621% and a Mean Average Recall of 97.279%. The model outperformed all other hand-crafted classifiers that were previ-

Figure 6. Samples of images of faces of elephants

Figure 7. Samples of images of elephants used for training the elephant recognition model.

Approach	MAP(%)	MAR(%)
LBP +KNN	28	82.353
LBP +SVM	80	57.97
LBP +Random Forest	86	78.12
SIFT + KNN	78	78.416
SIFT+ SVM	22	100
SIFT+ Random Forests	48	32.44
HOG+ KNN	32	82.353
HOG+ Random Forests	78	78.409
Dua .et.al[6]	100	63.29
Proposed Approach	**98.621**	**97.279**

Table 1. A comparison of various approaches applied for elephant recognition.

ously applied for detecting elephants.

4.4. Human Elephant Collision detection

The particle filter is used for tracking the movement of the elephants. Different scenarios for roads and railway tracks have been considered. The color of the particles changes from blue to red as soon as the elephant is detected in the region of high human intervention such as roads or railway tracks. It must be noted that, for detecting elephant intrusion on agricultural land, the complete area has been marked as an area of danger. As soon as an elephant is detected a warning is issued. Samples of instances of HEC on roads and railway tracks are shown in Figure 8 and Figure 9 respectively.

Approach	MAP	MAR
HOG+ KNN	72	63.158
HOG+ Random Forests	90	77.586
Dua .et.al[6]	72	63.158
Proposed Approach	**98.67**	**96.109**

Table 2. A comparison of various approaches applied for elephant face recognition.

Figure 8. A scenario of human-elephant collision on the road. The elephant is chasing a human being on a road. The particles appear to be in red as the elephant is continuously on the road.

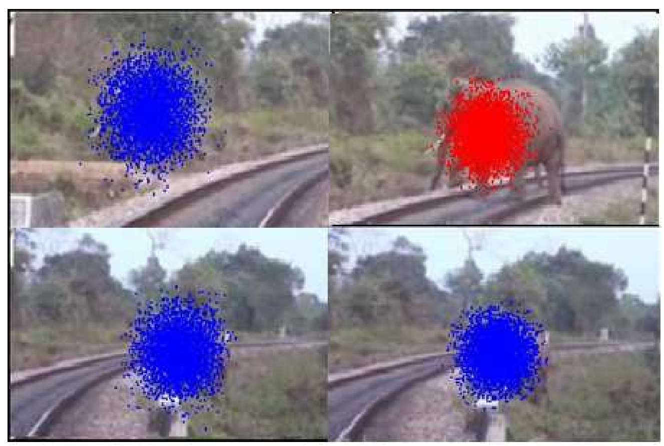

Figure 9. A scenario of depicting elephant intrusion on railway tracks. The elephant is crossing a railway track. The scenario can be dangerous for the elephant, if a train passes by. The particles detect an elephant and change the color as soon as the elephant crosses the area of conflict.

5. Conclusion

The paper proposes a framework to detect and prevent Human Elephant Collision. The various challenges involved with the development of a Human Elephant Collision detection system were taken into consideration and a robust framework was designed accordingly. The system was tested on various videos where HEC actually occurs. For a more precise differentiation of Elephants from other objects, two different recognition methodologies were applied to detect elephants. The results of the proposed techniques were compared with other techniques as well.

In future cases, the emphasis shall be on extending the work to real time scenarios. Hence, there is a need for an adequate system to warn the officials or the concerned authorities about human elephant collisions. Humans even at a farther distance from the place of collision should also be made aware in case the collision takes place. Infrared cameras can also be used to detect the presence of elephants at night.

References

[1] wild living asian elephant in India and south asia. `http://www.asianelephant.net/wildlife.htm`, 2011-2017. [Online; accessed 2011].

[2] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2274–2282, 2012.

[3] A. Ardovini, L. Cinque, and E. Sangineto. Identifying elephant photos by multi-curve matching. *Pattern Recognition*, 41(6):1867–1877, 2008.

[4] T. Burghardt and J. Ćalić. Analysing animal behaviour in wildlife videos using face detection and tracking. *IEE Proceedings-Vision, Image and Signal Processing*, 153(3):305–312, 2006.

[5] R. Dabarera and R. Rodrigo. Vision based elephant recognition for management and conservation. In *Information and Automation for Sustainability (ICIAFs), 2010 5th International Conference on*, pages 163–166. IEEE, 2010.

[6] I. Dua, P. Shukla, and A. Mittal. A vision based human-elepahant collision detection system. In *Image Information Processing (ICIIP), 2015 Third International Conference on*, pages 225–229. IEEE, 2015.

[7] R. Farah, J. P. Langlois, and G.-A. Bilodeau. Rat: Robust animal tracking. In *Robotic and Sensors Environments (ROSE), 2011 IEEE International Symposium on*, pages 65–70. IEEE, 2011.

[8] V. Goswami, M. Lauretta, M. Madhusudan, and K. Karanth. Optimizing individual identification and survey effort for photographic capture–recapture sampling of species with temporally variable morphological traits. *Animal Conservation*, 15(2):174–183, 2012.

[9] S. L. Hannuna, N. W. Campbell, and D. P. Gibson. Segmenting quadruped gait patterns from wildlife video. 2005.

[10] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1725–1732, 2014.

[11] L. E. King, A. Lawrence, I. Douglas-Hamilton, and F. Vollrath. Beehive fence deters crop-raiding elephants. *African Journal of Ecology*, 47(2):131–137, 2009.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[13] A. Mammeri, D. Zhou, A. Boukerche, and M. Almulla. An efficient animal detection system for smart cars using cascaded classifiers. In *Communications (ICC), 2014 IEEE International Conference on*, pages 1854–1859. IEEE, 2014.

[14] D. Ramanan and D. A. Forsyth. Using temporal coherence to build models of animals. In *null*, page 338. IEEE, 2003.

[15] P. Shukla, T. Gupta, A. Saini, P. Singh, and R. Balasubramanian. A deep learning frame-work for recognizing developmental disorders. In *Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on*, pages 705–714. IEEE, 2017.

[16] S. Sugumar and R. Jayaparvathy. An improved real time image detection system for elephant intrusion along the forest border areas. *The Scientific World Journal*, 2014, 2014.

[17] D. Tweed and A. Calway. Tracking multiple animals in wildlife footage. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 2, pages 24–27. IEEE, 2002.

[18] A. B. Venkataraman, R. Saandeep, N. Baskaran, M. Roy, A. Madhivanan, and R. Sukumar. Using satellite telemetry to mitigate elephant–human conflict: An experiment in northern west bengal, india. *Current Science*, pages 1827–1831, 2005.

[19] C. Vermeulen, P. Lejeune, J. Lisein, P. Sawadogo, and P. Bouché. Unmanned aerial survey of elephants. *PloS one*, 8(2):e54700, 2013.

[20] C. Wang. Video classification with convolutional neural network. *Proceedings of the 9th Ph. D. retreat of the HPI Research School on service-oriented systems engineering*, 100:245, 2017.