[1]*Parama Bagchi
[2]Olga S. Sushkova
[3]Alexei A. Morozov
[4]Debotosh Bhattacharjee

TCDLN: Terahertz Colormap Deep Learning Network for Hidden Object Detection

Journal of Electrical Systems

Abstract: - This paper is based on the recognition of hidden objects which is embedded in colormaps generated by using special programs in terahertz video surveillance systems. Terahertz imaging is popular because of its capability to see through opaque objects. So, this imaging technique can be used in the detection of hidden objects, medical diagnosis, and many other real-life applications. This paper has two significant contributions: firstly, how to detect hidden objects in colormaps, and secondly to diagnose which colormap gives the best detection results. We shall involve a deep learning framework that utilizes Resnet-50, to detect hidden objects concealed within the color maps of terahertz images. Secondly, we have proposed a method for inspecting purple, grey, inverted cool, and grey+skeleton colormaps using a deep learning framework and concluded which colormap detects the hidden object most robustly.

Keywords: Colormaps, Deep Learning, Terahertz, Security System

1. INTRODUCTION

Terahertz imaging is significant because it enhances any organization's security system. They are advantageous because they can help detect malicious persons who have hidden dangerous objects under their clothing. The terahertz video surveillance dataset we use in this work is a new dataset [1]. The novelty of our dataset is that it is combined with a 3D video that allows one to understand what parts of the terahertz image belong to the human body, and we have extracted these parts automatically using a program written in Actor Prolog. Although researchers worldwide have already created several versions of the terahertz video imaging dataset, specific works on object detection using the terahertz imaging dataset are still to be explored. This paper has a clear goal. We have generated special diagrams (colored images) from the terahertz images generated by the terahertz video datasets and used them to detect hidden dangerous objects. For detection, we have used Resnet-50 as a tool and chalked out a hypothesis as to the objects of which colormaps the neural network detects the best. Also, we have tried to find out which colored image the neural network failed to identify and why.

2. MAIN OBJECTIVES

Terahertz imaging is still not used for real-life applications due to the following main reasons:
1. Terahertz cameras for data acquisition are too costly.
2. It is also difficult to carry terahertz cameras for capturing outdoor subjects/objects because they are bulky and too slow, making them impractical for real-life applications. This is because modern terahertz cameras use mechanical scanning systems (with rotating metallic mirrors, etc.). Everything will change when a new generation of terahertz cameras is created based on matrix terahertz detectors.
The main problem is that manual inspections of people are tougher methods for preventing terrorist attacks. Let us consider a manual inspection method. In the case of a conference, there are huge gatherings of people all around the place where the conference is being organized. Suppose we employ manual inspection; huge crowds will be stuck in a
queue because the person in charge will have to check each person individually. Manual inspection allows people to be checked inside a conference room where a social event is organized. We have to check everyone who is coming inside the conference room. We have to control all the doors of the conference room. The main problem is that a huge crowd will be created because of the restrictions on people's movement. So, if terrorist attacks occur, many people will fall prey to them.

[1]*Corresponding author: Parama Bagchi, Dept of CSE, RCC Institute of Information Technology, Kolkata, India
paramabagchi@gmail.com
[2] Olga S. Sushkova, Kotelnikov Institute of RadioEngineering and Electronics of RAS, Mokhovaya 11-7, Moscow 125009,Russia
o.sushkova@mail.ru
[3] Alexei A. Morozov, Kotelnikov Institute of RadioEngineering and Electronics of RAS, Mokhovaya 11-7, Moscow 125009,Russia
morozov@cplire.ru
[4]Debotosh Bhattacharjee, Dept of CSE, Jadavpur University, Kolkata, India
debotosh@ieee.org

A possible solution is using terahertz video surveillance for the preliminary fast check of the people. This video surveillance system could help to detect terrorists far away from the place where they are being manually checked. Let us consider an outdoor scenario. Mostly during festivals, there are huge gatherings. In this case, a terahertz video surveillance system will also enable us to detect terrorists from far away. “Far away” is from where the density of the crowd is high, thus ensuring the safety and security of the general public as a whole. Fig 1 shows the manual inspection by a vigilance officer at an ongoing conference. It is clear from Fig-1 that the manual inspection is weak enough to control large crowds who have come together at an event. So, a surveillance system should be employed since this is a question of public security. Fig 2 shows the automatic inspection using the terahertz camera to manage a large crowd and determine who the terrorist is. Now, the question is, can the intruder be located from far off only using a terahertz camera? The answer is that, three policemen need to control the large crowd assembled to attend a conference. The policeman with a stick controls people’s entry inside the hall. The policeman with the terahertz camera watches if any suspicious person is in the crowd. The third policeman is standing at one side of the hall with the ‘safe’ sequence of 36 people to be allowed inside the conference room after a manual check.

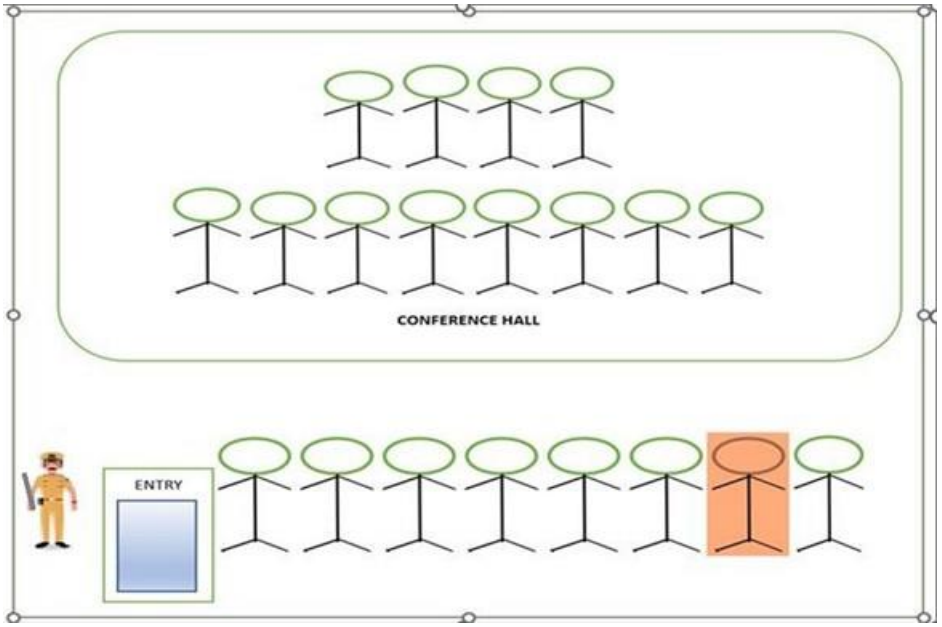


Fig . 1. Manual inspection is practically impossible for tracking who the terrorist is. The manual inspection system cannot handle large crowds of people.

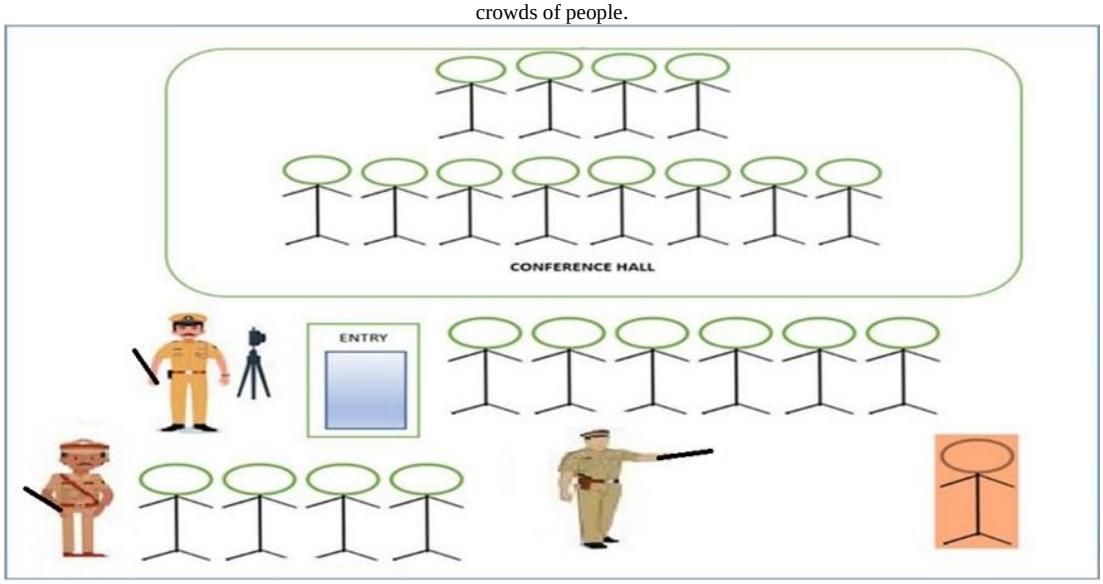


Fig. 2. The terahertz inspection can supplement the manual inspection and safe people from sudden attacks during the manual inspection*

2.1 Proposal To Solve The Problems Of Terahertz Imaging

In this paper, our major contribution is to solve the problem of detecting dangerous hidden objects using a terahertz video surveillance system. Manual detection is not impossible, but quite cumbersome to detect a malicious person who is hiding any dangerous object under his/her clothing. Also, using this system, a person

may not allow him/her to be checked manually. This is a problem existing in many countries. But this problem could be well tackled using a terahertz video surveillance camera. We propose to capture the images of people in the real-time system using this camera. We plan to employ the terahertz camera to capture live images of individuals. The only possibility to detect an intruder in the system is using this camera.

3. RELATED WORKS

Terahertz (THz) technology is now a widespread application, particularly due to its ability to detect hidden objects. It is used not only for object detection but also in the biomedical domain. The advantage of terahertz image capturing is that it is fast enough and ensures accurate capturing technology. First, let us discuss a brief history of terahertz imaging [2, 3]:

- In 1960, terahertz imaging was first generated.
- In 1995, the THz image was first generated using time-domain spectroscopy.
- In 2002, the first terahertz of a hand was generated by ESA(European Space Agency)
- In 2004, ThruVision Ltd demonstrated the first THz camera successfully capturing guns hidden under clothing.
- In mid-2007, scientists at the U.S. Department of Energy's Argonne National Laboratory, and collaborators in Turkey and Japan, announced the creation of a device that can lead to a portable battery source of T waves. Ulrich Welp of Argonne's Materials Science Division led this group.
- In 2008, at Harvard University several hundred nanowatts of terahertz radiation was emitted.
- In 2009, it was proved that T waves are not polarized. The observed spectrum peaked at 2 THz and a broader peak was at 18 THz. The radiation was not polarized.
- In 2011, Rohm and a research team at Osaka University produced a terahertz emission of 1.5Gbit/s.
- In 2013, Georgia Institute of Technology's Broadband Wireless Networking Laboratory and the Polytechnic University of Catalonia developed graphene antenna that would broadcast terahertz frequencies. The range of frequencies was 10 to 100 nanometers wide and one micrometer long.

Several experiments on terahertz imaging have been conducted by the research team at IRE RAS [4], where they have used a terahertz camera to design a video surveillance system. In one of their works [5], they created methods and software tools for the automatic analysis of video images in the terahertz frequency range. Several notable works in the development of superconducting integrated receivers for radio astronomy and atmospheric monitoring have already been done(http://www.cplire.ru/html/lab234/publications.htm), and the research group at IRE RAS has contributed to terahertz imaging [6] domain.

This section presents a survey on hidden object detection in the field of terahertz imaging. Fig 3 shows the various types of terahertz imaging techniques.

Terahertz imaging can be of three types namely active, passive, and hybrid. The works performed in these three categories can be subdivided into four broad categories in which maximum work has been done. These are:

(a) Harmless object detection
(b) Multi-object resolution-based object detection
(c) Deep learning-based object detection and
(d) Mining-based object detection.

One of the limitations is that the hybrid model is not fully explored. Nowadays, hybrid models are being proposed for very fast acquisition of terahertz imaging, but how they will perform to detect hidden objects is still unknown.

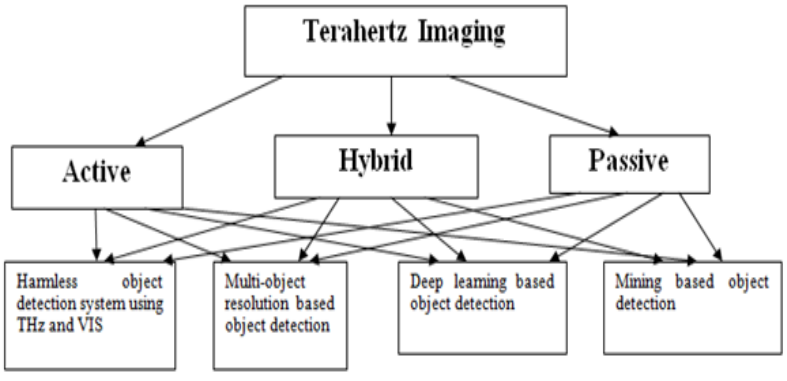


Fig. 3. Diagram depicting the categories of terahertz imaging.

The main difference between active and passive methods is that an active system radiates electromagnetic waves that are considered safe for the people but nobody knows whether using these devices in real life is safe or not. This is a similar problem with smartphones. The main disadvantage of an active system is that, they do not support long-distance traversal due to their absorption by atmospheric water vapor. Passive systems also have their disadvantages. For instance, they cannot detect objects if not in front of the human body. Also, passive systems do not possess the large dynamic range of active systems.

Ref (Author, year)	Active/passive/hybrid imaging	Details of dataset	Detection rate	
Kowalski et. al.[7], 2013	Active	Fusion of THz and VIS Images	NA	
Liang et al.[8], 2021	Active	11 classes of objects	RetinaNet	91.46%
			YOLOv4	84.49%
			FRCN-OHEM	83.86%
Xiao et al.[9], 2018	Active	The SUN data set contains 397 class scenarios. The MS COCO dataset has 6 to 7 targets per image.	Faster R-CNN	77.3%
			R-FCN	78.4%
			Optimized Faster R-CNN	79.4%
			R-PCNN	84.5%
Zhang et al.[10], 2018	Active	5 sets of objects	Using CNN, an accuracy of 96.98% was achieved.	
Li et al.[11], 2021	Active	1347 hidden objects	RetinaNet	91.46%
Qi et al.[12], 2015	Active	A 330 GHz active terahertz imaging system was developed.	NA	NA
Stroescu et al.[13], 2019	Active	2442 different images of the six roadside targets	The highest accuracy of 98.78% on the testing dataset was achieved.	
Danso et. al [14], 2022	Active	329 images	The YOLOv5-fpn detector recorded the highest accuracy of 99.4%.	
Kowalski et. al. [15], 2015	Passive	Three types of objects were used: a plastic pistol, a metal knife, and one type of clothing.	The highest accuracy of 57.13% was attained using a technique based on the normalized pixel intensity of the object.	
Kowalski et. al. [16], 2019	Passive	Various objects were used: plastic pistols, metal pistols, bombs, ceramic knives, metal knives, leather wallets, and mobile phones.	The highest accuracy of 94% was attained using R-FCN.	
Garcia et al. [17], 2018	Active + Passive	Images obtained from each sensor were	The final image provided a reduced cost compared to custom-made	

		combined using image fusion	equipment's.

Table I. Literature Review

The category 'a' deal with object detection systems using THz and VIS cameras. Category 'b' deal with multi-object-based object detection. Category 'c' deal with deep learning-based object detection. Finally, category 'd' deal with data mining techniques for object detection.

In the next section, we shall discuss each category 'a-d' one by one and their related works already performed. Let us first talk about active terahertz image acquisition. In [7], the author aimed to build a harmless system for screening individuals to detect hidden objects using THz and VIS cameras. The authors undertook various quantitative analyses based on a fusion of THz and VIS cameras using the SSIM index of THz fused images. The best result for the SSIM THz fused images using a ratio pyramid was 0.0347. The key feature of this system was that only terahertz images were not analyzed, but a fusion of terahertz and VIS images was analyzed.

Table 1 shows a literature review of the existing work done in the field of terahertz imaging. Liang et al. [8] had captured a public dataset using an active terahertz imaging camera with a 5 by 5 mm resolution. There are four males and six females in this dataset. A total of 1349 objects in various modes have been captured. Experimental results show that the detection rate (DR) of RetinaNet was 91.46%. But the reason why some network detects some objects better is not explained here, which was a shortcoming.

Xiao et al. [9] proposed a framework (R-PCNN) that combined pre-processing and a faster version of R-CNN. The maximum detection rate was 84.5% and the detection time was less than 20 milliseconds. In [10], an improved detection method based on deep learning for terahertz image detection was invented. Based on that, the Terahertz Human Dataset and the Classification Dataset were established. Then, considering the similarity between terahertz and optical images, the authors proposed the classification method based on transfer learning of optical features. Here, the authors also proposed a segmentation method based on bounding slipping. The highest detection rate was 96.98%.

Li et al. [11] devised a dataset for multi-object detection in active terahertz imaging. But, due to poor imaging quality, object detection was more difficult than public object detection datasets already available. The authors used YOLOv3, YOLOv4, FRCN-OHEM, and RetinaNet on this dataset and experiments proved that RetinaNet achieved has the best detection rate.

Qi et al. [12] deduced a 330 GHz active terahertz imaging system for hidden object detection. In this system, the rotation speed and the angle have been designed so that, using Delaunay triangulation, colormaps, and grey-scale images have finally been obtained. Stroescu et al. [13], devised another active imaging application, where six different roadside targets in low THz radar were set up using convolutional neural networks. These were employed for radar imagery classification. This application was used for all weather sensing systems. Here, the highest accuracy of 98.78% on the testing dataset was achieved. However, the reason for the analysis of the colormaps is not known.

In [14], the authors developed an effective active imaging system where BiFPN and YOLOv5 were used as deep learning tools. In this method, a maximum detection rate of 99.6% was recorded.

Next, let us discuss passive image detection in the case of terahertz imaging. Kowalski et al. [15] worked on passive image detection based on metal objects covered with clothing. To investigate the time stability of detection, they performed measurements in sessions, each lasting for 30 mins. They have presented a theoretical comparison of two spectra and the results of experiments. The innovation in this work was that the author combined properties of imagers and radiation in the spectral ranges. This method achieved the highest detection rate of 57.13% using a detection based on Normalized pixel intensity. In [16], the authors presented another passive imaging system for the detection and recognition of concealed objects with various types of clothing by using passive imagers operating in a terahertz (THz) range at 1.2 mm (250 GHz) and an infrared (MWIR) wavelength of 3–6 μm (50–100 THz). During this study, a large dataset of images covered with various types of clothing has been collected. Those are implemented on the NVIDIA 1080-Ti graphical processing unit and are reported to process a maximum of 14 and 7 frames per second for YOLO3 and R-FCN, respectively. Another work was proposed in [17], where an active cum passive system using a depth-sensing camera was evolved. Here the system was simulated at the hardware and software levels. The resulting images of the sensors are combined using image fusion techniques, using an optical sensor to remove noise, highlighting the object in a subject's body. In [18], a millimeter-wave imaging system was developed, especially for indoor applications. The authors showed that this system offers better visibility for concealed objects. The results demonstrate that multispectral illumination can reveal objects hidden by artifacts. The third form of imaging is hybrid imaging, which utilizes active and passive THz imaging to retrieve full information about the object. We use a passive THz camera, and our THz data was collected using the THERZ-7A industrial passive terahertz video surveillance system (Astrohn Technology Ltd) [19].

With respect to the existing works, in the field of terahertz imaging, we have proposed a novel approach to investigate different colormaps encapsulating hidden objects for Terahertz video surveillance using Deep learning using the statistical analysis of colormaps, which the above works have not dealt with. The major drawback of all the above systems was that each had devised new deep-learning frameworks to detect hidden objects. But, none of them have devised reasons why the neural network failed to detect certain objects and successfully detected others. This is an entirely new contribution that we have addressed here. In this paper, we have written special programs that output various colormaps obtained when terahertz images from our terahertz video surveillance system were given as input. These colormaps were used to detect dangerous hidden objects, which could be used to analyze real-time surveillance systems. This is probably an entirely new idea that has never been used before.

4. Important Contributions

Our main contributions have been enlisted as follows.

• First, we've experimented with four different colormaps embedding several objects, namely the purple, grey, inverted cool, and grey with skeleton colormaps generated by the passive THz camera, and have thus compared which colormap detects the hidden objects better.

• Our main contribution is using special diagrams to review why some colormaps have given higher object detection results than others.

• We have devised an economical technique to insert checkpoints to coach our network for a substantial time. This contribution is useful for people using online services to train their systems. Inserting checkpoints saves time and cost and is also useful for somebody who wants to train their systems for long hours. Most applications need substantial time to train their systems to get precise performance.

Also, we have used several interesting image augmentation techniques to take care of the variability of our dataset, although the number of frames is huge in our present dataset.

In this work, we propose a new scenario of using terahertz video surveillance. Unfortunately, we cannot experiment in the outdoor scene and show that it works, but we have shown how it can be used to analyze the colormaps of terahertz imaging for hidden object detection.

5. Materials And Strategies

This section discusses the acquisition of the THz dataset used for this work. The dataset contains videos of individuals with harmful and non-harmful objects hidden under their garments. A special platform is developed to accumulate and pre-process the video data [20] of our present terahertz system. The coding platform includes a translator of the Actor Prolog programming language to Java.

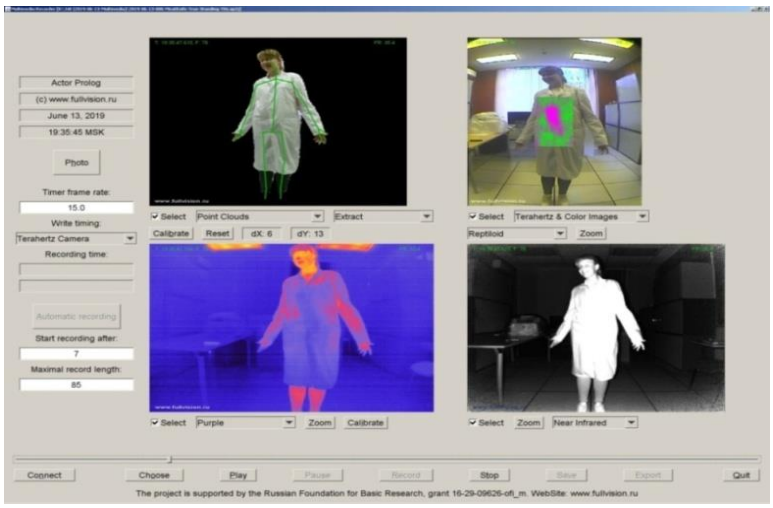
The idea used in the paper is that we have researched the various colormaps, and then deduced which colormap is appropriate for object detection. In this respect, we have provided a mathematical analysis of the colormaps showing how the various objects are to be viewed using the various colormaps. Thus, we have provided a statistical description of the various colormaps and concluded that which colormap provides the best detection rate.

• Stages of creation:

In this section, we will discuss the steps of THz colormap dataset preparation. At first, the frames are extracted from the THz video surveillance dataset, which is given as input to the system. The assorted colormaps, namely the purple, grey, inverted cool, and grey with skeleton colormaps, are used from the terahertz images. These terahertz images are generated using special programs. Every frame contains the following information:

1. THERZ-7A produces an integer matrix of size 110x180 with each element of 2 bytes. The operating spectral range of the camera is 0.23–0.27 THz.
2. The present system produces an RGB image with a resolution of 576×704.
3. A 3D array, namely a point cloud, is produced by the camera of the Kinect v2 device. The point cloud is a matrix that has a resolution of 512×424.
4. RGB image is produced by Kinect v2 and consequently projected onto the 3D image. Thus, the 1920×1080 initial resolution of Kinect v2 RGB images is decreased to 512×424 to save memory.
5. A near-infrared image is produced by Kinect v2. The resolution is 512×424.
6. Also, the images of 3D skeletons of people are detected by Kinect v2. 2D and 3D frames of point clouds of persons are detected in the videos.
7. The thermal images of resolution 640×480 that the TE V1 thermal camera has produced are also detected.

The data has been published on the website [20]. The data consists of video files of various persons, as shown in Fig 4. Fig 4 shows a subject with a harmful object hidden under her clothing. The figures in the middle show a subject with four different figure modes, the top left shows a 3D range image, the top right shows a terahertz, the bottom left shows a thermal image, and the bottom right shows an infrared image. However, the near-infrared and thermal images are not used in our experiment.



(a)

(b) (c)

Fig. 4. (a)Top-most diagram depicting the image acquisition process pertinent to terahertz imaging(top-left figure demonstrates a 3D image, top-right figure demonstrates a terahertz image with the object detected, bottom-left figure demonstrates a thermal image and bottom-right figure demonstrates a near-infrared image)(b) Terahertz image of a subject with the object detected (in pink color) (c) Terahertz image of the subject across a specific pose with object and skeletons detected

We have used an additional software package called the Actor Prolog and an open-source library along with it. We downloaded it from the developer's website [20]. Fig 5 shows the user interface of the Multimedia_3D_normalizer program, which was used to prepare the experimental dataset.

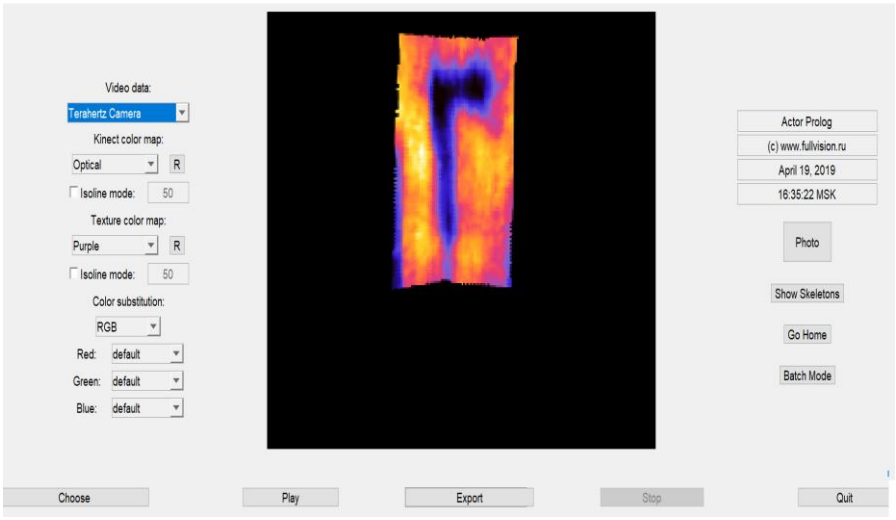


Fig. 5. GUI for the Multimedia_3D_normalizer program showing the object termed as 'AXE'

The Multimedia_3D_normalizer is a program of multimodal data preparation and experimentation with machine learning tools. The idea is that several images are combined into one. A semantic fusion of the images is done here, which is important for successful training using any machine learning tools. In the schema of the video fusion, the coordinates of the person's skeleton are recognized by the Kinect-2 device and those are used for correcting cloud position. Fig 5 shows the user interface when we executed this program on the Actor Prolog software platform. In Fig 5, we can see that there is a select button for selecting the terahertz video dataset. There are several buttons for selecting the "texture colormaps". When the user clicks on the "Choose" button, a taskbar comes up to choose the video dataset, and the "play" button plays the video dataset, the "export" button exports the video dataset into a drive, the "stop" button stops playing the video dataset.
After the Actor Prolog software had generated the colored images using the selected colormaps, we downloaded and saved them in our drive. To detect the objects in our colormaps, we have implemented the deep learning framework Resnet-50 [21] on this dataset. The architecture of Resnet-50 contains an extra layer added as the last layer. The last layer contains 2048 nodes. Since our number of classes is seven, the output layer contains 7 nodes. The problem is that the video file contains very similar frames when they are close to one another. Therefore, the data set's variability is quite low despite the number of frames being big. Thus, we use data augmentation methods to increase the variability of the training data set. The following augmentation techniques have been applied both to the training and the test datasets namely:

- Horizontal Flip.
- Autocontrast.
- AdjustSharpness.
- Gaussian Blur.
- RandomRotation about 15°.

5.1 Developing real-life terahertz images

The above acquisition techniques were developed to be used under controlled situations. But real-life applications of terahertz imaging need to be studied in detail. First, while developing a real-life application using terahertz imaging, we need to capture images of people under uncontrolled conditions like festivals, conferences, and symposiums where gathering people is subsequently large in number. These are highly challenging jobs and demanding because we could easily trace the terrorists. The terahertz images can be much more complex than analyzing images acquired in our labs. There are many reasons for this i.e., some people use medical bondage because of problems with their health. These people will look like terrorists with bombs in a terahertz image. This fact must be considered during the preliminary inspection of people using terahertz cameras. Another problem is that wet clothes do not penetrate the terahertz waves. Therefore, the terahertz camera will not be able to detect dangerous objects during the rain or immediately after the rain because the clothes are wet. People can use dangerous objects that were not considered in our experiments. They can place these objects in non-standard places not considered in our experiments.
Moreover, passive terahertz imaging cannot detect objects placed in bags and suitcases. We hope our results will stimulate progress toward real-life analysis of terahertz videos because our works will be optimized to tackle real-world video surveillance problems in the future. We have tried our best to use augmentation techniques very close to real life. Indeed, we have not encountered problems with wet clothes, because more experiments are necessary in the future.

6. Experimental Results And Discussion

Now, we shall demonstrate our major contributions in separate sections.

6.1. Data Preparation

Our main goal of the experiment is to check which colormaps are the most appropriate for the computer analysis of the THz video data that contains images of dangerous and prohibited objects. To demonstrate the validity of our approach, we have utilized four colormaps for training and testing a Resnet-50 network. These four colormaps are:

· Grey colormap
· Purple colormap
· Grey colormap with skeleton
· Inverted cool colormap

The size of each of these datasets for the above colormaps is 328 x 7= 2296 because there are 328 instances of seven dangerous objects in each colormap. When these colored terahertz images are fed to the neural network, we have redefined a procedure that guarantees that the same random number vector separates the dataset to the training and validation datasets for different colormaps.

6.2 Network Training

After data preparation is over, the next stage is the training of the network. In this case, the same indices for the training and validation datasets for all the colormaps were applied, and the images were saved. Initially, these four colormaps are fed to the deep neural network. For training purposes, the colormaps were fed to our deep learning framework. In this case, we used the Adam optimizer, with a learning rate of '1e-5', and the weight_decay was set to 0. The colored images were computed using the colormaps. We have trained the neural network for 400 epochs using checkpoints. The experimental results associated with these colormaps are enlisted in Fig 6.

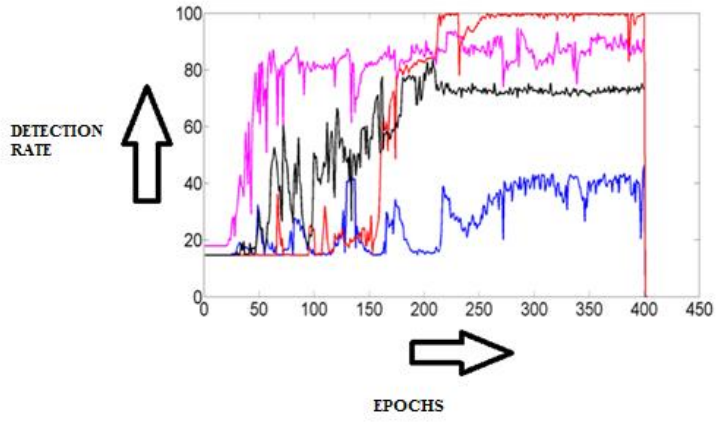
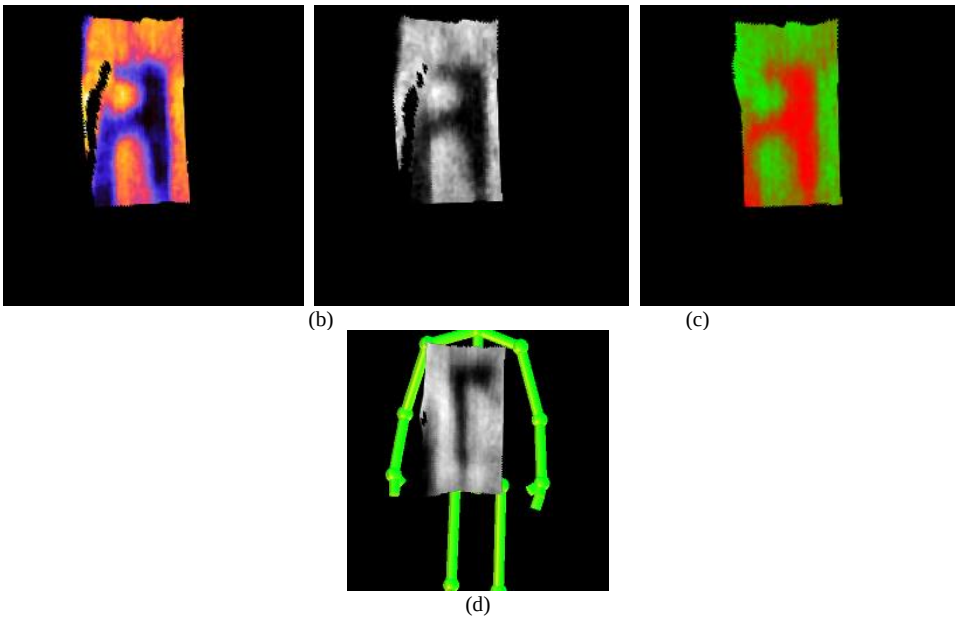


Fig. 6. Detection rates by Resnet 50 (magenta color - Purple colormap (about 95%), black color – Grey colormap (about 75%), blue color - Grey+Skeleton colormap (about 40%), red color - Inverted Cool colormap (about 100%))

Using the Resnet-50 network, we achieved about 95% detection rate in the case of the purple colormap and 100% in the case of the inverted cool colormap, whereas a detection rate of about 75% was obtained in the case of the grey colormap and the lowest detection rate of about 40% was obtained in case of grey with skeleton colormap. The dynamics of the training process depict that, in the case of the purple colormap, the detection rate increased steadily and attained a recognition rate of 95%. In the case of the grey colormap, initially, the neural network's detection rate increased and then slowly decreased, which shows that the neural network is slow in detection. In the case of the grey+skeleton colormap, the detection rate is the lowest as it shows that the neural network is having great difficulty detecting the objects. For the "inverted cool" colormap, the neural network was slow initially but attained a 100% detection rate after 200 epochs. To explain the above results, we have created red, green, and blue diagrams, as shown in Fig 8, corresponding to each colormap, as shown in Fig 7.



(a) (b) (c)

(d)

Fig. 7. Examples of terahertz images visualized using the colormaps: Purple colormap, Grey colormap, Inverted Cool colormap, and grey+skeleton colormap.

The algorithm for creating these special diagrams is depicted in Algorithm 1.

Algorithm 1

Step 1: Select the colormap to be analyzed.

Step 2: Read the image and input the R, G, and B channels in separate vectors.

Step 3: The first figure will indicate the pixels that correspond to the range 0-0.5 of the intensity in the given color channel.

Step 4: The third figure will indicate the pixels that correspond to the range 0.5-1 of the intensity in the given color channel.

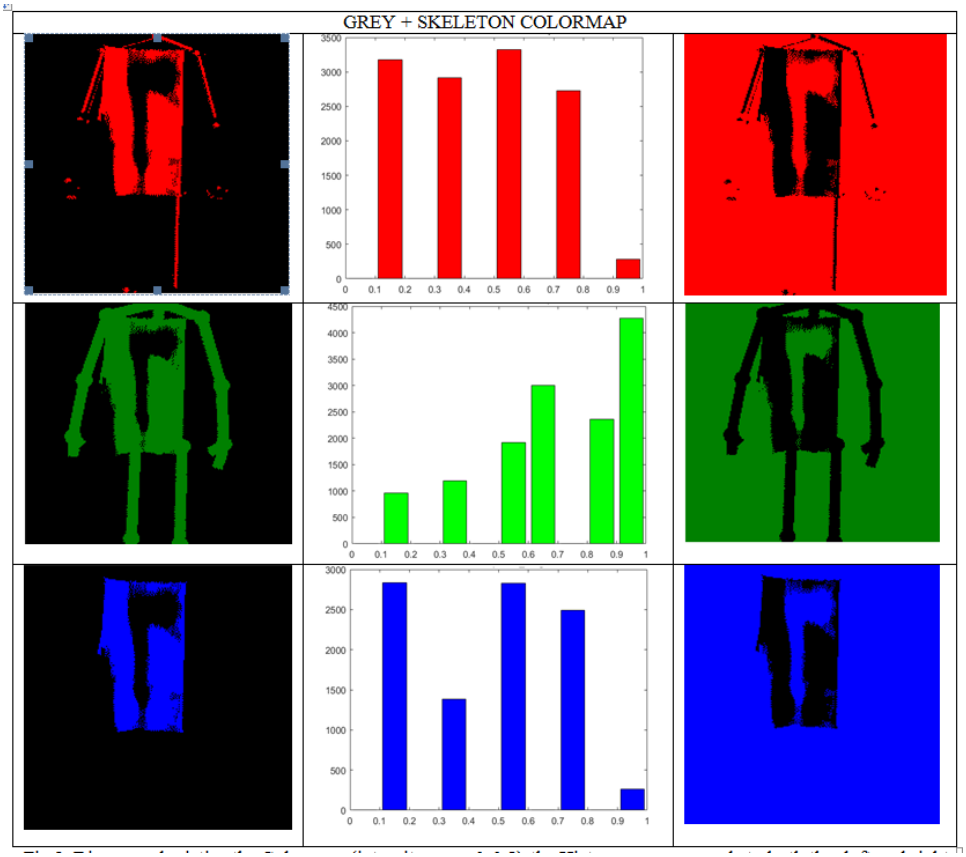


Fig. 8. Diagrams depicting the Colormaps (intensity range 0-0.5), the Histogram corresponds to both the left and right colormaps, and the colormaps (0.5-1 intensity range).

The confusion matrices in 'Appendix A' show the actual values vs. the predicted values for each colormap. We have seven classes of objects numbered 0 to 6. For the first colormap, the number of object '0' instances are 166. And it is truly detected as object 0(indicated by the number in the diagonal elements). The correctly detected object is depicted in the diagonal elements. The incorrectly detected object (false positive) is detected in a bold color in the same row as non-diagonal elements, e.g. in case of the first object '0' the purple colormap and the grey+skeleton colormap have detected many instances of object '0' incorrectly. Grey and inverted cool colormaps have shown far better performance in this case because the deep learning framework has detected all instances of the first object correctly. Finally, the worst performance is shown by the grey plus skeleton colormap because except objects '0' and '1', all other objects are falsely detected. And the best performance is recorded by an inverted cool colormap because almost all objects are correctly detected, as shown by the diagonal elements.

6.3 Data Augmentation

For data preparation, we have used the image augmentation methods discussed above.
After augmentation, the images look as shown in Fig 9.

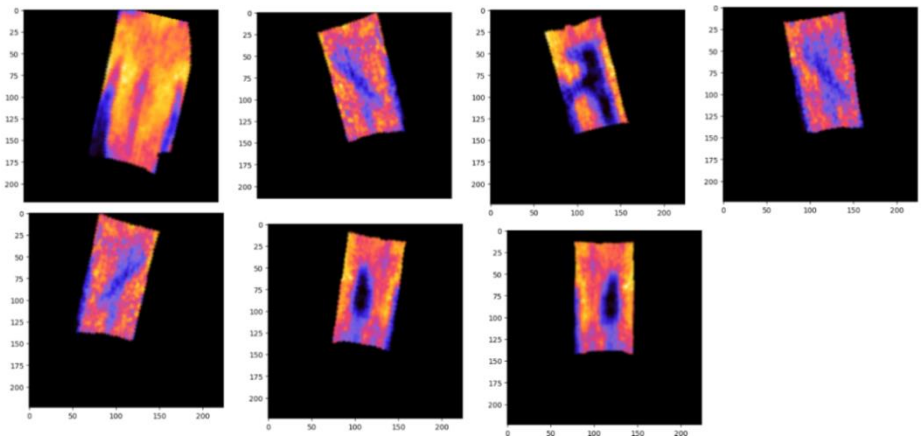


Fig. 9. Final augmented images for the purple colormap.

6.4 Checkpointing Issues

We have introduced the checkpoints to train the system and obtain favorable outputs. Checkpoints allow us to train the system for a substantial period. We have used the following algorithm termed as 'save_checkpoint' for saving the trained model using the checkpoints.

Algorithm save_checkpoint

Step 1: Declare the state of the model in a file with the extension '.pth'
Step 2: Now save the present state in the file '.pth'
End

From the main body of the program, we need to call the save_checkpoint function as:

If (epochs == 100)
Call save_checkpoint (model.state_dict(),<file_path>)

After 100 epochs, if the user wants to load the state, he will simply call:
Call model.load_state_dict (<file_path>)

6.5 Analysis of the results using the colormaps and diagrams

Our last contribution is that we have used special diagrams described above to prove why some colormaps detect the objects much better than others. As an illustration, Fig 10 shows the various Red, Green, and Blue channels we have obtained for each colormap, namely the grey colormap, inverted cool colormap, grey+skeleton colormap, and purple colormap.
Now, we shall analyze why the neural network detects a colormap better than the others. Let us consider the inverted cool colormap first. Here, the red channel contains a pure image of the object, and the green channel contains an image of the object and the rectangle. Now, it will be very clear to somebody if he looks at the red channel of the inverted cool colormap, which only shows the target object without any other geometric figure, i.e., the rectangular border. The above logic holds for the purple colormap, also. The blue channel contains the object and has no rectangular border.
Alternatively, in the case of the grey colormap, all three channels, namely red, green, and blue, contain the target object along with an additional geometric figure, so the neural network's detection rate is lower than the above two. Additionally, in the case of the grey and skeleton-based colormap, the detection rate is the lowest. During the training process, we input some images to our deep learning framework. These images contain the target and, in some cases, rectangles and other objects. Therefore, we train the neural network to recognize the target object and ignore all the other objects. It is difficult for the neural network, to understand what we want to recognize when all images contain both target objects and rectangles. That is why the rectangles influence the

training process too much. Additionally, one of the reasons why the inverted cool colormap is better detected than the purple colormap is that most of the objects in the green channel of the purple colormap contain large holes, as shown in the last row of Figure 10, even though the best part is there are no added rectangles. This is one of the reasons why the inverted cool colormap detects images better than the purple colormap.

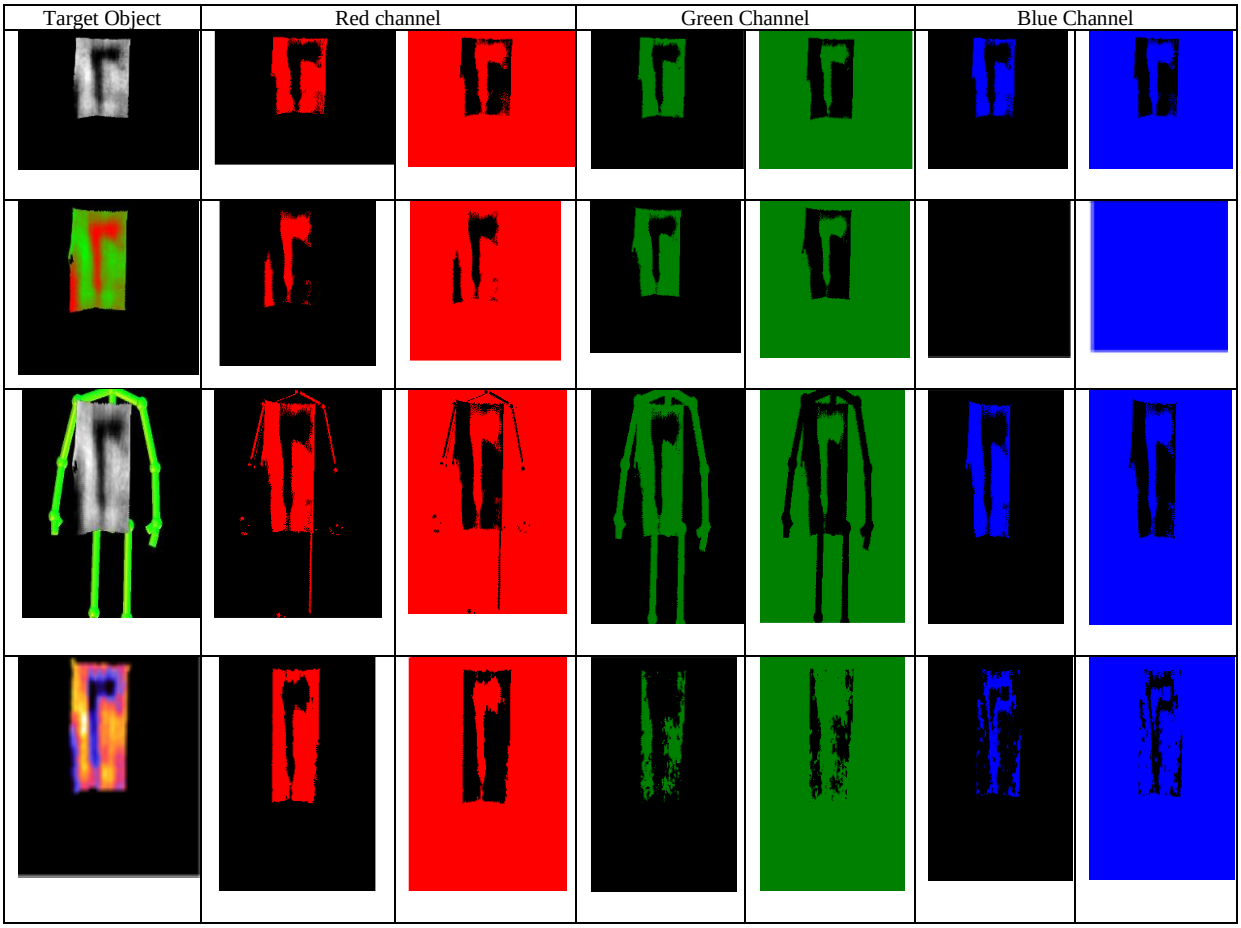


Fig. 10. Analysis of the colormaps pertaining to the target objects (Grey colormap, inverted cool colormap, grey+skeleton colormap, and purple colormap).

7. Comparative Analysis

In this section, we shall draw a brief comparison between the existing works performed on passive terahertz imaging and our present work. Here, we shall analyze how our present system outperforms the others. One of the main differences between the present work and existing works is that we have used different colormaps for the subjects. These colormaps will either enhance or degrade the appearance of the objects. We have a new gap that we have analyzed, which colormap could give better results in such cases, using deep learning framework. The main difference between our work and existing works is that we have addressed different problem statements. For instance, in [22], the researchers used the SSD algorithm; in [23], the YOLO3 and R-FCN algorithms were used; and in [24] the YOLOV5 detector was used. In [25], the researchers have used YOLO-MSFG in various object scenes and attained higher detection rates.

Maximum people distinguish the target objects using various deep learning techniques, but we have detected them without detecting their coordinates. In the present case, other researchers have tried to detect the coordinates of target objects (using the YOLO architecture). But our main idea was to analyze the colored images computed using the colormaps. Now, we shall discuss each of the above existing works individually. In [22], two objects, namely one dangerous and a non-dangerous object were detected using the SSD algorithm. The work resulted in a detection rate of 99.94% in the case of the object gun and 99.9% in the case of the object phone. In [23], the authors detected two objects, namely a gun and a knife, using the R-FCN algorithm resulting in a detection rate of 93% in the case of a knife and 92% in a gun.

Data augmentation methods were used in this work. The authors in [24], used terahertz imaging for digital holography using YOLOv5 detector. The authors in [25] used YOLO-MSFG on various object scenes, and the highest accuracies obtained were 93.49% and 90.76% on single and multi-class objects. Here, two augmentation techniques, namely rotate and flip, were used. However, the authors in [26] worked on 3D object detection and reconstruction. Also, no augmentation techniques were used.
Our present work surpassed all the previous works because, we have investigated which colormaps are better for solving the hidden object detection problem and why. In this work, Resnet-50 has been initialized with the deeply trained features of the colormaps and attained a maximum recognition rate of 100%. In this work, we have experimented with colormaps which embed hidden objects. We used four colormaps, namely the purple, grey, grey+skeleton, and inverted cool colormap. The objects we have used have all been enlisted in Appendix 'A.'

8. Conclusion And Future Scope

Our main idea in this paper was to investigate the objects of which colormaps is best detected by the neural network and why. In addition, we investigated which objects of the colormaps the neural network fails to identify and why. We have supported our hypothesis by the use of special kinds of diagrams developed to analyze the colormaps, which have helped us to explain the experimental results described in the paper. On one hand, when the highest detection rate of 100% is given by the inverted cool colormap, this fact supports our hypothesis; on the other hand, the bad results demonstrated by the Grey with skeleton dataset indicate that standard neural network architectures like Resnet50 cannot utilize additional information in the form of the skeletons to improve the detection rate; therefore special network architectures should be developed to utilize this additional information during the terahertz video surveillance. This can be a prospective research area in the future.

Author Biographies

Parama Bagchi received her B. Tech degree in CSE from BCET (Under WBUT) in 2005, M. Tech degree in Computer Technology from Jadavpur University in 2010 and Ph.D degree in CSE from Jadavpur University in Dec 2021. She is the recipient of the University Gold medal in 2010 from Jadavpur University for getting the highest marks in M.TECH. She worked in different private engineering colleges and management institute in Kolkata initially as lecturer (CSE), visiting faculty, and then in the post of Assistant Professor (CSE) until 2014. At present, she is posted as Assistant Professor (CSE) (Senior Scale, Grade-11), at RCC Institute of Information Technology, Beliaghata, Kolkata. She is an ACM professional member, and has been the reviewer of many prestigious journals and conferences. She is currently associated with three collaborative projects with Kotel'nikov Institute of Radio Engineering and Electronics of Russian Academy of Sciences, Moscow, Russia, on Terahertz Video surveillance system, with Tula State University, Russia on Biomedical Engineering, and the Institute of Computational Science, University of Potsdam, Germany and University of Kragujevac, Serbia on Smart Farming applications, and Jadavpur University. Her current research interests pertain to 3D objects, smart farming, terahertz video surveillance system and applications of machine learning in medical domain.

Olga S. Sushkova defended her Ph.D. thesis at Bauman Moscow State Technical University in 2017. She is a senior research associate of Kotel'nikov Institute of Radio Engineering and Electronics of Russian Academy of Sciences. She is a 2019 Moscow Government Award laureate. Her research interests include biomedical signal analysis, Parkinson's disease, statistical analysis of biomedical signals, and development of methods for differential diagnosing of neurodegenerative diseases.

Alexei A. Morozov graduated from Bauman Moscow State Technical University in 1991. In 1998, he defended his Ph.D. thesis on logic programming. He is a senior research associate of Kotel'nikov Institute of Radio Engineering and Electronics of Russian Academy of Sciences. His research interests include object-oriented logic programming, intelligent video surveillance, terahertz video surveillance, and biomedical signal analysis (EEG, EMG, MEG). He is a member of the International Association for Logic Programming (ALP). He is a principal developer of the Actor Prolog language (www.fullvision.ru).

Debotosh Bhattacharjee (SM'11) received M.C.S.E. and Ph.D. (Engineering) degrees from Jadavpur University, India, in 1997 and 2004, respectively. He was associated with different institutes in various capacities until 2007. He then returned to his alma mater, Jadavpur University. He has authored or coauthored over 304 journal and conference publications, and several book chapters in the area of biometrics. During his postdoctoral research, he visited several universities in Europe, including the University of Twente, The Netherlands, Instituto Superior Técnico, Lisbon, Portugal, and Heidelberg University, Germany. He is a Life Member of the Indian Society for Technical Education, New Delhi, and the Indian Unit for Pattern Recognition and Artificial Intelligence. His research interests pertain to the applications of computational intelligence techniques, such as fuzzy logic, artificial neural networks, genetic algorithms, and rough set theory in face recognition, OCR, and information security.

ACKNOWLEDGEMENT

Authors are grateful to the Kotelnikov Institute of Radioengineering and Electronics of Russian Academy of Sciences, Mokhovaya, Moscow, Russia (http://www.cplire.ru/html/index.html) for its support of our research work.

Appendix A List of target objects used in our experiments

Axe	2019-04-19-003-Axe-True-Standing-THz	
Bottle	2019-04-19-012-Bottle (Vodka)-True-Standing-THz	
AK	2019-04-19-019-AK-True-Standing-THz	
MeatKnife	2019-06-13-006-MeatKnife-True-Standing-THz	
TT	2019-06-13-008-TT-True-Standing-THz	
ShoulderHolster	2019-06-13-029-ShoulderHolster (TT)-True-Standing THz	
Tin	2019-06-17-010-Tin-True-Standing-THz	

Appendix B URLs of the data downloaded from the developer's website

URL OF WEBSITE: https://www.fullvision.ru/monitoring/description_eng.php

- 2019-04-19-003-Axe-True-Standing-THz (https://disk.yandex.com/d/ehn1E7v8lU5vhg)

- 2019-04-19-012-Bottle (Vodka)-True-Standing-THz (https://disk.yandex.com/d/C39UI2NGHg9AwA)
- 2019-04-19-019-AK-True-Standing-THz(https://disk.yandex.com/d/XR4ibG0FJyDauA)
- 2019-06-13-006-MeatKnife-True-Standing-THz(https://disk.yandex.com/d/tivKUP6_pnITQA)
- 2019-06-13-008-TT-True-Standing-THz(https://disk.yandex.com/d/t-gXR5r94ldhLQ)
- 2019-06-13-029-ShoulderHolster (TT)-True-Standing-THz (https://disk.yandex.com/d/Td1oNTqaOJmaeQ)
- 2019-06-17-010-Tin-True-Standing-THz (https://disk.yandex.com/d/9CTBrHsNqyM7yQ)

Appendix C Test Confusion Matrices for all the datasets

Inverted cool colormap

	Actual Values						
Predicted Values	166	0	0	0	0	0	0
	0	175	0	0	0	0	0
	0	0	146	0	0	0	0
	0	0	0	165	0	0	0
	0	0	0	0	158	0	0
	0	0	0	0	0	164	0
	0	0	0	0	0	0	168

Grey+skeleton colormap

	Actual Values						
Predicted Values	142	0	0	0	0	0	24
	0	144	0	0	0	0	31
	0	0	0	0	0	0	146
	0	0	0	0	0	0	161
	0	0	0	0	0	0	163
	0	0	0	0	0	0	164
	0	0	0	0	0	0	168

Purple colormap

	Actual Values						
Predicted Values	77	0	91	0	0	0	0
	0	108	0	2	0	67	0
	0	0	147	0	0	0	0
	0	0	2	164	0	0	0

	0	0	2	0	163	0	0
	0	0	0	0	0	163	0
	0	0	9	6	0	0	199

Grey colormap

	Actual Values						
Predicted Values	166	0	0	0	0	0	0
	0	0	0	0	0	175	0
	0	0	38	0	0	108	0
	0	0	0	166	0	0	0
	0	0	0	1	159	3	0
	0	0	0	0	0	164	0
	1	1	0	0	0	20	146

References

[1] A. A. Morozov, and O. S. Sushkova, "Development of a publicly available terahertz video dataset and a software platform for experimenting with the intelligent terahertz visual surveillance," *Proceedings of International Conference on Frontiers in Computing and Systems: COMSYS 2020*, pp. 105-113 (2021).

[2] P. K. Sethy, P. R. Mishra, and S. Behera, "An introduction to terahertz technology, its history, properties and application," *Proc. International conference on computing and communication*, (2015).

[3] F. F. Sizov, "Brief history of THz and IR technologies," S*emiconductor physics, quantum electronics & optoelectronics 22, No 1: 67-79* (2019).

[4] A. Morozov, O. Sushkova, A. Polupanov, V. Antsiperov, G. Mansurov, S. Paprotskiy, A. Gorchakov, A. Yanushko, N. Petrova, and A. Bugaev, "A Method of Terahertz Intelligent Video Surveillance Based on the Semantic Fusion of Terahertz and 3D Video," (2019).

[5] A. A. Morozov, O.S. Sushkova, A.F. Polupanov, G.K. Mansurov, S.K. Paprotskiy, A.V. Yanushko, N.G. Petrova, A.S. Bugaev, and V.E. Antsiperov, "Semantic fusion and joint analysis of terahertz and 3D video images by the means of object-oriented logic programming," *Radioelectronics. Nanosystems. Information Technologies*, 10(2), pp.329-340 (2018).

[6] V. P. Koshelets, P. N. Dmitriev, A. B. Ermakov, L.V. Filippenko, A.V. Khudchenko, N.V. Kinev, O.S. Kiselev, I.L. Lapitskaya, A.S. Sobolev, M.Y. Torgashin, and P.A. Yagoubov, "Superconducting Integrated Receivers for Radio Astronomy and Atmospheric Monitoring," (2007).

[7] M. Kowalski, N. Palka, M. Piszczek, and M. Szustakowski, "Hidden object detection system based on fusion of THz and VIS images," *Acta Physica Polonica A 124*, no. 3: 490-493 (2013).

[8] D. Liang, F. Xue, and L. Li, "Active terahertz imaging dataset for concealed object detection," *arXiv preprint arXiv:2105.03677*, (2021).

[9] H. Xiao, R. Zhang, H. Wang, F. Zhu, C. Zhang, H. Dai, and Y. Zhou, " R-pcnn method to rapidly detect objects on thz images in human body security checks, " *IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)* ,pp. 1777-1782, IEEE (2018).

[10] J. Zhang, W. Xing, M. Xing, and G. Sun, "Terahertz image detection with the improved faster region-based convolutional neural network," *Sensors, 18(7)*, p.2327 (2018).

[11] L. Li, F. Xue, D. Liang, and X. Chen, "A Hard Example Mining Approach for Concealed Multi-Object Detection of Active Terahertz Image," *Applied Sciences*, 11(23), p.11241 (2021).

[12] C. C. Qi, G. S. Wu, Q. Ding, Y. D. Zhang, and B.D. Xixiang, "A 330 GHz active terahertz imaging system for hidden objects detection".

[13] A. Stroescu, M. Cherniakov, and M. Gashinova, "Classification of high resolution automotive radar imagery for autonomous driving based on deep neural networks," *20th International Radar Symposium (IRS), pp. 1-10*, IEEE, (2019).

[14] S.A. Danso, L. Shang, D. Hu, J. Odoom, Q. Liu, and B. Nana Esi Nyarko, " Hidden dangerous object recognition in terahertz images using deep learning methods," *Applied Sciences*, 12(15), p.7354 (2022).
[15] M. Kowalski, and M. Kastek, "Comparative studies of passive imaging in terahertz and mid-wavelength infrared ranges for object detection," *IEEE Transactions on Information Forensics and Security*, 11(9), pp.2028-2035 (2016).
[16] M. Kowalski, "Hidden object detection and recognition in passive terahertz and mid-wavelength infrared," *Journal of Infrared, Millimeter, and Terahertz Waves 40*, no. 11-12: 1074-1091 (2019).
[17] F. Garcia-Rial, D. Montesano, I. Gomez, C. Callejero, F. Bazus, and J. Grajal, "Combining commercially available active and passive sensors into a millimeter-wave imager for concealed weapon detection," *IEEE Transactions on Microwave Theory and Techniques*, 67(3), pp.1167-1183.
[18] L. Zhang, J. Stiens, A. Elhawil, and R. Vounckx, "Multispectral illumination and image processing techniques for active millimeter-wave concealed object detection," Applied *optics, 47(34), pp.6357-6365* (2008).
[19] http://astrohn.com
[20] https://www.fullvision.ru/
[21] H. A. Khan, W. Jue, M. Mushtaq, and M.U. Mushtaq, "Brain tumor classification in MRI image using convolutional neural network," *Mathematical Biosciences and Engineering*, (2021).
[22] L. Cheng, Y. Ji, C. Li, X. Liu, and G. Fang, "Improved SSD network for fast concealed object detection and recognition in passive terahertz security images," *Scientific Reports*, 12(1), p.12082, (2022).
[23] M. Kowalski, "Hidden object detection and recognition in passive terahertz and mid-wavelength infrared," *Journal of Infrared, Millimeter, and Terahertz Waves*, no. 11-12: 1074-1091, (2019).
[24] K. Cheng, and Q. Li, "Application of acquiring region of interest based on the YOLOv5 model in terahertz digital holography, *Applied Optics*, no. 14, 3589-3597, (2023).
[25] F. Xu, X. Huang, Q. Wu, X. Zhang, Z. Shang, and Y. Zhang, "YOLO-MSFG: toward real-time detection of concealed objects in passive terahertz images, " *IEEE Sensors Journal*, 22(1), pp.520-534, (2021).
[26] A. Mathai, N. Guo, D. Liu, and X. Wang, "3D transparent object detection and reconstruction based on passive mode single-pixel imaging," *Sensors*, no. 15: 4211 (2020).
[27] *https://stock.adobe.com/in/search?k=%22indian+police%22&asset_id=559092967