CEUR-WS.org/Vol-2391/paper19.pdf

Development of a Method of Terahertz Intelligent Video Surveillance Based on the Semantic Fusion of Terahertz and 3D Video Images

A A Morozov[1], O S Sushkova[1], I A Kershner[1] and A F Polupanov[1]

[1]Kotel'nikov Institute of Radio Engineering and Electronics of RAS, Mokhovaya 11-7, Moscow, Russia, 125009

e-mail: morozov@cplire.ru, o.sushkova@mail.ru, ivan kershner@mail.ru

Abstract. The terahertz video surveillance opens up new unique opportunities in the field of security in public places, as it allows to detect and thus to prevent usage of hidden weapons and other dangerous items. Although the first generation of terahertz video surveillance systems has already been created and is available on the security systems market, it has not yet found wide application. The main reason for this is in that the existing methods for analyzing terahertz images are not capable of providing hidden and fully-automatic recognition of weapons and other dangerous objects and can only be used under the control of a specially trained operator. As a result, the terahertz video surveillance appears to be more expensive and less efficient in comparison with the standard approach based on the organizing security perimeters and manual inspection of the visitors. In the paper, the problem of the development of a method of automatic analysis of the terahertz video images is considered. As a basis for this method, it is proposed to use the semantic fusion of video images obtained using different physical principles, the idea of which is in that the semantic content of one video image is used to control the processing and analysis of another video image. For example, the information about 3D coordinates of the body, arms, and legs of a person can be used for analysis and proper interpretation of color areas observed on a terahertz video image. Special means of the object-oriented logic programming are developed for the implementation of the semantic fusion of the video data, including special built-in classes of the Actor Prolog logic language for acquisition, processing, and analysis of video data in the visible, infrared, and terahertz ranges as well as 3D video data.

1. Introduction

Recently, the terahertz range of the electromagnetic waves attracts a strong interest of the safety systems developers [1–9]. This interest is caused by a set of special properties of the terahertz radiation. For instance, the terahertz radiation can penetrate dielectric materials like plastic, wood, and ceramics. The terahertz radiation is safe for people and can be used in public places in contrast with the X-radiation. Furthermore, the terahertz range of the electromagnetic waves includes the resonance frequencies of complex molecules and, therefore, the terahertz spectroscopy can be used for the distant detection of explosives, drugs, and other dangerous substances.

The terahertz range of the electromagnetic waves is situated between the microwaves and the infrared radiation (see figure 1). It is accepted that the frequency of the terahertz radiation is about 3 THz – 300 GHz that corresponds to the wavelengths from 0.1 to 1 millimeter. Actually, the bounds of the terahertz range are conventional; they are defined differently in research papers.

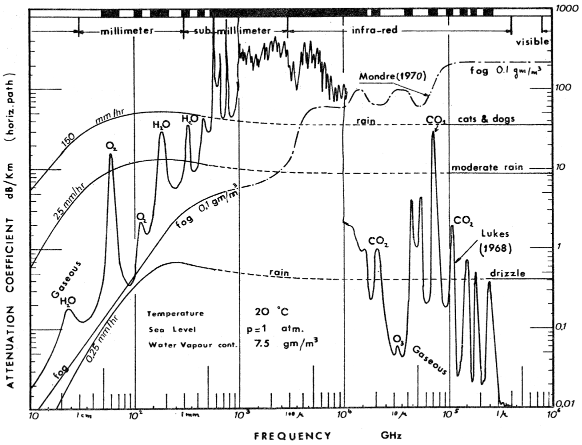


Figure 1. This is the dependence of the attenuation coefficient of the electromagnetic waves penetrated in the atmosphere on the wavelength of the radiation [10]. The terahertz waves are situated between the microwaves and the infrared radiation and correspond approximately to the area of high values of the attenuation coefficient. The abscissa is the frequency [THz] and the ordinate is the attenuation coefficient [dB/Km].

It is significant that the properties of the terahertz radiation and the principles of its usage differ for various sub-ranges of the terahertz waves. In particular, the 0.5-3 THz waves are used for the implementation of the terahertz spectroscopy and detection of dangerous substances [11]. Detection of the weapons and other dangerous objects hidden under the clothing of people is usually based on the usage of terahertz radiation frequencies that are less than 1 THz (so-called sub-terahertz radiation) that correspond to the transparency windows of the clothing. Active, passive, and combined methods of the sounding are used for the detection of the hidden objects.

There is a substantial difference between the images of hidden objects acquired using the active and passive sounding methods. Accordingly, the analysis of the terahertz images of different kinds also requires solving different problems and application of different methods.

The passive terahertz video surveillance is based on the receiving the essential human body radiation. In this case, the extrinsic objects look like dark spots against the background of the intrinsic emission of the human body (see an example in figure 2). Main problems of the passive terahertz image processing are the following ones:

(i) Typical passive terahertz images are fuzzy and unclear. The resolution of the images and the signal-to-noise ratio are low.

(ii) The background of the typical passive terahertz image is dark in comparison with the human body image. The shades of the hidden objects look like dark areas too. Therefore, any mistake in the separation of the foreground and background in the terahertz images automatically leads to the erroneous detection of hidden objects and false alarms.

The active terahertz video surveillance requires a target illumination and the registration of the radiation reflected from the human body (see an example in figure 3). The problems of active terahertz image processing are mostly caused by the fact that the reflection of the external terahertz radiation sources produces flares of different kinds. These flares have often prolonged

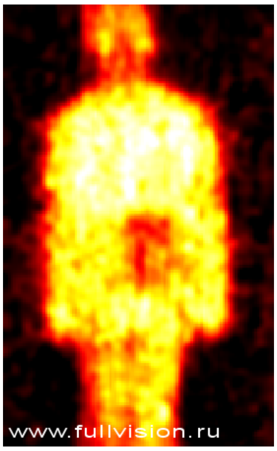


Figure 2. This is a human body image in the terahertz range. The image is acquired using the THERZ-7A industrial passive terahertz video surveillance system (Astrohn Technology Ltd [12]). The frequency range is 0.23–0.27 THz. The TT gun is hidden behind the belt at the back.

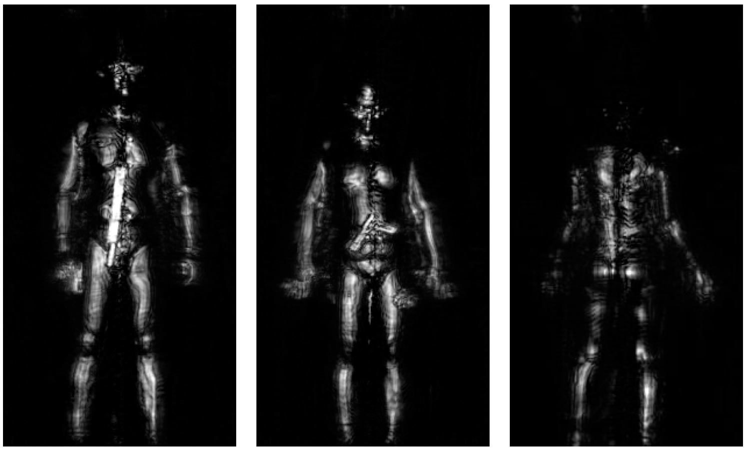


Figure 3. These are examples of human terahertz images taken with an active terahertz video surveillance system [13]. In the figure, there are samples of cold weapon and fire-arms hidden under the clothing of the persons.

shapes that can be mistakenly recognized as a cold weapon or other dangerous objects hidden under the clothing [13].

At present, the main directions of the terahertz video surveillance development are the combination of the active and passive methods of terahertz video acquisition and implementation of 3D terahertz video surveillance. In particular, these problems were addressed recently in the framework of the CONSORTIS European project [14, 15] (see an example in figure 4). Unfortunately, there is still no evidence of the development of fully-automatic hidden objects detection methods that are reliable enough to be used in the industrial terahertz video surveillance systems.

2. Semantic fusion of heterogeneous video images

Fundamentally different methods are necessary for the implementation of the fully-automatic analysis of terahertz video images and recognition of hidden objects. It is necessary to take into account the semantics of the video images including the context of the video scene on the

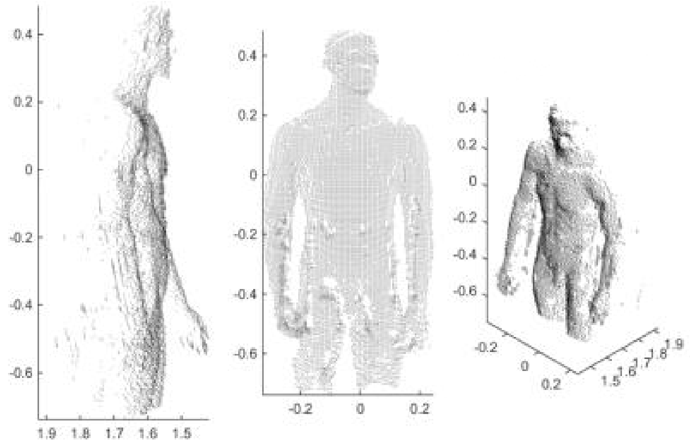


Figure 4. This is an example of 3D terahertz image of a mannequin acquired using the Pathfinder 200 GHz terahertz radar [14].

analogy of how the human operator analyzes the terahertz images. The additional information that is to be taken into consideration includes the coordinates of the body, arms, and legs of the person, multi-spectral video information (video, infrared, terahertz, etc.), time variations of these attributes, etc. The consideration of this information is especially important when the terahertz video surveillance system has to watch the free movements of the persons in a public place. Next, we will call the fully-automatic and semi-automatic terahertz video surveillance systems as terahertz intelligent video surveillance systems by analogy with the conventional intelligent video surveillance systems that operate in the video and/or infrared spectral ranges.

A typical terahertz video image looks like a set of fuzzy spots that can be monochromatic or colored depending on the data analysis method applied. A conventional terahertz video surveillance system displays a video in the visual and/or infrared range simultaneously with the terahertz video. This video information enables to the specially trained operator to interpret the terahertz image in a proper way and to detect objects hidden under the clothing of the visitors. This work of the human operator is a kind of semantic fusion of heterogeneous video images. The idea of the semantic fusion is in that several videos are to be united so that the semantic content of one video image is used to control the processing and analysis of another video image.

It is the authors' opinion that one of the most important data sources for the object recognition in the terahertz video is the positional relationships between the body, arms, and legs of the person and the terahertz video image. It is advisable to use a point clouds and the images of skeletons of the persons acquired by a time-of-flight camera for this purpose. To implement this idea, a set of special built-in classes of the Actor Prolog object-oriented logic language [16–27] were developed: *Astrohn*, *KinectBuffer*, *TEV1*, etc.

The *Astrohn* built-in class implements the terahertz and RGB video data acquisition using the THERZ-7A device [12]. The *Astrohn* class supports the data input from the device as well as reading from and writing to the video file. The *Astrohn* class supports conversion of the terahertz video data to the color video images. In particular, pseudo colors can be used for the terahertz data representation. The *Astrohn* class operates also with RGB video acquired from the internal IP-camera of the THERZ-7A device and can combine this RGB video data with the terahertz video. The *Astrohn* class implements a simple synchronization of the terahertz and RGB video streams. For this purpose, each terahertz frame is coupled with the RGB frame that is the nearest in time. Currently, the *Astrohn* class supports more than 25 high resolution color maps including a set of conventional thermal imaging color maps: Aqua, Blackhot, Blaze, BlueRed, Gray, Hot, HSV, Iron, Red (Jet), Medical, Parula, Purple, Reptiloid, and Green (Rainbow).

The *KinectBuffer* built-in class acquires 3D video data from the time-of-flight camera of the Kinect 2 device (Microsoft Inc). The reading and recording of 3D video data files are also supported [28]. The following essential functions are implemented in the *KinectBuffer* class:

(i) Creation of the 3D surface based on the 3D point cloud.

(ii) Projection of given texture to the surface using a 3D lookup table [26, 28, 29].

We have used these features of the *KinectBuffer* class to the fusion of 3D and terahertz video images in our experiments. A special method of the speculative reading of the video files is implemented in the *Astrohn* class that enables to synchronize recorded 3D and terahertz video data.

An example of a 3D image that is generated by the fusion of a time-of-flight camera point cloud and a terahertz image is demonstrated in figure 5. The terahertz video is combined with the image of a person's skeleton that was computed by the procedures of the standard Kinect 2 SDK. A 3D lookup table was applied to project the terahertz video to the 3D surface in the real time. In particular, the user can rotate, zoom, and shift the 3D video by the mouse during the demonstration. In the example, the 3D point cloud is recognized as a human body and this information is used for the selection of terahertz image colored areas that are directly related to the objects hidden under the clothing of the person. This is a case of semantic fusion of heterogeneous video information that prevents false detections of background terahertz areas as target hidden objects.

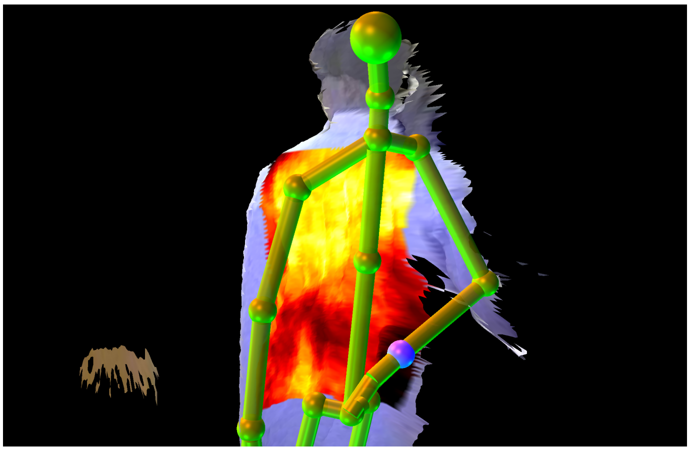


Figure 5. This is an example of 3D and terahertz video data fusion implemented using the *KinectBuffer* and *Astrohn* built-in classes of the Actor Prolog language [24, 28].

3. An example of heterogeneous video data analysis

Let us consider an example of heterogeneous video data analysis. The goal of this experiment is to check whether the terahertz videos contain information enough to teach a convolutional network to distinguish the dangerous and safe objects.

A set of heterogeneous videos was prepared for the experiment (see figures 6 and 7). For that, a special logic program was written in Actor Prolog for multichannel video data acquisition (see figure 8). The video includes 3D point clouds and terahertz images of persons. A calibration procedure [29] was performed to compute a 3D lookup table that establishes relations between the video images of different kinds. Then another logic program was written to project terahertz images to the 3D images of the persons and to generate training/test data sets in the PNG

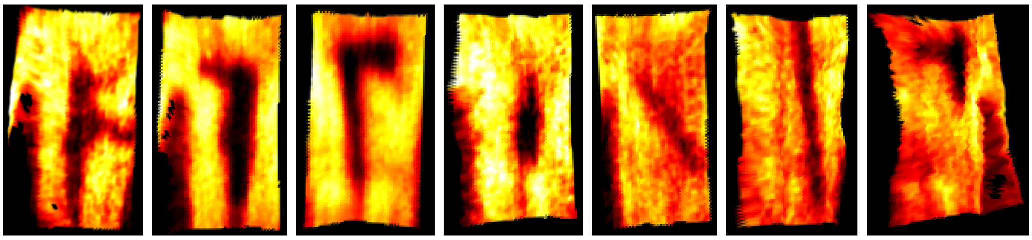


Figure 6. The learning set includes weapon and other dangerous objects (from left to right): the Kalashnikov sub-machine-gun (AK), AK without the magazine, an axe, bottles, a knife, a baton, and guns of different brands.

format. An image generated by this logic program is shown in figure 5. The difference between the image 5 and the images demonstrated in figures 6 and 7 is in that the later images were rotated and normalized to provide the uniform size and angle of view for all frames. Besides, the images of skeletons and RGB video data were eliminated. The frames with inappropriate positions of the person in the view area were automatically discarded. The Hot standard color map was used for the terahertz data visualization.

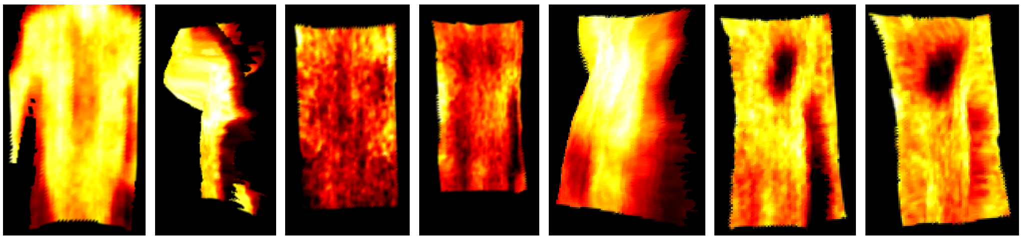


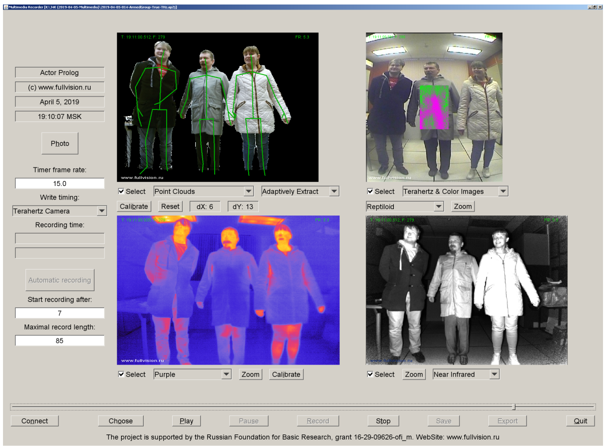
Figure 7. The learning set includes also terahertz images of people dressed in casual clothes and outer clothing. Some images contain ordinary objects like phones and USB disks. The number of these images is balanced with the number of images that contain weapon and dangerous objects.

Convolutional networks of several standard architectures were trained using the data sets: LeNet [30], AlexNet [31], ResNet50 [32], and Darknet19 [33]. The results of the training are reported in table 1. It is not a surprise that the oldest network LeNet yields the worst results and the Darknet19 that is the latest of these four networks yields the best results.

After that, an additional test data set was prepared that includes only the images of a person that keeps the M16 automatic rifle and the images of the person without extra objects (see figure 9). The number of images of different kinds was balanced. Then, the trained networks were used to analyze the video images.

The results of the experiment are reported in table 2. The networks recognize successfully the M16 automatic rifle as a dangerous object. Surprisingly, the AlexNet architecture yields the best results in spite of the fact that this network architecture is quite old and simple. The newest Darknet19 architecture yields unexpectedly the worst results in this test. Probably this is because the recognition and the generalization are different problems and the development of network architectures for the generalization of video data requires a special attention.

These results demonstrate that the neural network approach to the terahertz video data analysis can make generalizations of the hidden object properties and successfully predict that the hidden object is a kind of a weapon and/or dangerous object. It is a promising area for further research to make experiments with heterogeneous video data fusion, standardizing of



Multimedia Recorder [K:\34K (2019-04-05-Multimedia)\2019-04-05-014-ArmedGroup-True-THz.ap2j]

Actor Prolog
(c) www.fullvision.ru
April 5, 2019
19:10:07 MSK

Timer frame rate: 15.0
Write timing: Terahertz Camera
Recording time:
Start recording after: 7
Maximal record length: 85

☑ Select Point Clouds Adaptively Extract
dX: 6 dY: 13
☑ Select Terahertz & Color Images
Reptiloid
☑ Select Purple
☑ Select Near Infrared

The project is supported by the Russian Foundation for Basic Research, grant 16-29-09626-ofi_m. WebSite: www.fullvision.ru

Figure 8. This is the user interface of the logic program written in Actor Prolog for the automation of multichannel video data acquisition. The program controls the video data acquisition simultaneously from the THERZ-7A device (the right window on top), the Kinect 2 device (the left window on top and the bottom right window), and the i3system TE V1 thermal camera (the bottom left window). The video data is recorded in a special Actor Prolog format developed for the multichannel video data processing.

Table 1. These are the results of the training of the convolutional networks of various architectures. The size of the training set is 9173 video frames. The size of the test data set is 2293 frames. The training process includes two stages: 30 epochs (that is 5520 iterations) without transformations and 30 epochs with the flip and warp transformations. The image size is 224×224. The batch size is 50.

Network	Accuracy	Precision	Recall	F1 Score
LeNet	0.8203	0.8224	0.8203	0.8272
AlexNet	0.8543	0.8545	0.8543	0.8558
ResNet50	0.9930	0.9931	0.9930	0.9930
Darknet19	0.9974	0.9974	0.9974	0.9974

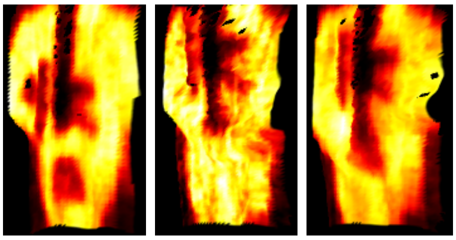


Figure 9. This is terahertz images of a person that keeps the M16 automatic rifle.

video data by non-linear color maps, and development of neural network architectures for the terahertz data analysis.

Table 2. The trained networks recognize the M16 automatic rifle as a dangerous object with the following quality. The size of the test data set is 672 video frames.

Network	Accuracy	Precision	Recall	F1 Score
LeNet	0.8720	0.8981	0.8720	0.8865
AlexNet	0.9970	0.9970	0.9970	0.9970
ResNet50	0.9940	0.9941	0.9940	0.9941
Darknet19	0.7589	0.8373	0.7589	0.8058

4. Conclusion

A method of semantic fusion of heterogeneous video data is proposed as a basis for the implementation of the terahertz intelligent video surveillance. In the framework of this method, 3D video data is used for the analysis and proper interpretation of the terahertz videos. Special logic programming means were developed for the experimenting with the terahertz video surveillance including a set of built-in classes of the Actor Prolog language for terahertz, infrared, and RGB video data acquisition, writing, reading, and synchronization. It was demonstrated that these logical means enable real-time video data acquisition and processing. In particular, the terahertz video can be projected to the 3D human body surface acquired by a time-of-flight camera. This heterogeneous information can be used by video data analysis algorithms to establish the positional relationships between the body, arms, and legs of the person and the colored areas in the terahertz video that helps to improve the detection of the objects hidden under the clothing of the person.

5. References

[1] Federici J F, Schulkin B, Huang F, Gary D, Barat R, Oliveira F and Zimdars D 2005 *Semiconductor Science and Technology* **20** S266

[2] Chan W L, Deibel J and Mittleman D M 2007 *Reports on progress in physics* **70** 1325

[3] Sanders-Reed J N 2015 *Micro- and Nanotechnology Sensors, Systems, and Applications VII* (International Society for Optics and Photonics) **9467** 94672E

[4] Antsiperov V E 2016 Automatic target recognition algorithm for low-count terahertz images *Computer Optics* **40(5)** 746-751 DOI: 10.18287/2412-6179-2016-40-5-746-751

[5] Sizov F 2017 *Semiconductor Physics, Quantum Electronics & Optoelectronics* **20** 273-283

[6] Appleby R, Robertson D A and Wikner D 2017 P*assive and Active Millimeter-Wave Imaging XX* (International Society for Optics and Photonics) **10189** 1018902

[7] Dhillon S S, Vitiello M S, Linfield E H, Davies A G, Hoffmann M C, Booske J, Paoloni C, Gensch M, Weightman P, Williams G P, Castro-Camus E, Cumming D R S, Simoens F, Escorcia-Carranza I, Grant J, Lucyszyn S, Kuwata-Gonokami M, Konishi K, Koch M, Schmuttenmaer C A, Cocker T L, Huber R, Markelz A G, Taylor Z D, Wallace V P, Zeitler J A, Sibik J, Korter T M, Ellison B, Rea S, Goldsmith P, Cooper K B, Appleby R, Pardo D, Huggard P G, Krozer V, Shams H, Fice M, Renaud C, Seeds A, Stohr A, Naftaly M, Ridler N, Clarke R, Cunningham J E and Johnston M B 2017 *Journal of Physics D: Applied Physics* **50** 043001

[8] Chen S, Luo C, Wang H, Deng B, Cheng Y and Zhuang Z 2018 *Sensors* (Basel, Switzerland) **18** 1342

[9] Yuan J and Guo C 2018 *Eighth International Conference on Information Science and Technology (ICIST)* 159-164

[10] Zufferey C H 1972 A Study of Rain Effects on Electromagnetic Waves in the 1-600 GHz Range Master's thesis *The MIMICAD Research Center*

[11] Baker C, Lo T, Tribe W, Cole B, Hogbin M and Kemp M 2007 *Proceedings of the IEEE* **95** 1559-1565

[12] ASTROHN Technology Ltd 2019 URL: http://astrohn.com

[13] Zhang J, Xing W, Xing M and Sun G 2018 *Sensors* **18** 2327

[14] CONSORTIS 2018 *Final Publishable Summary Report* (Teknologian Tutkimuskeskus VTT)

[15] Robertson D A, Macfarlane D G and Bryllert T 2016 *Passive and Active Millimeter-Wave Imaging XIX* 9830 983009

[16] Morozov A A 1999 *IDL* (Paris, France) 39-53

[17] Morozov A A, Vaish A, Polupanov A F, Antciperov V E, Lychkov I I, Alfimtsev A N and Deviatkov V V 2014 *Biodevices Scitepress* 53-62

[18] Morozov A A and Polupanov A F 2014 *CICLOPS-WLPE* (Aachener Informatik Berichte no AIB) 31-45

[19] Morozov A A, Sushkova O S and Polupanov A F 2015 *RuleML DC and Challenge* (Berlin: CEUR)

[20] Morozov A A 2015 *Pattern Recognition and Image Analysis* **25** 481-492

[21] Morozov A A and Sushkova O S 2016 Real-time analysis of video by means of the Actor Prolog language *Computer Optics* **40(6)** 947-957 DOI: 10.18287/2412-6179-2016-40-6-947-957

[22] Morozov A A, Sushkova O S and Polupanov A F 2017 *Advances in Soft Computing* (Cham: Springer International Publishing) **II** 42-53

[23] Morozov A A, Sushkova O S and Polupanov A F 2017 *ISIE* (Washington: IEEE Xplore Digital Library) 1631-1636

[24] Morozov A A, Sushkova O S and Polupanov A F 2019 *Optoelectronics in Machine Vision-Based Theories and Applications* (IGI Global Publications) 134-187

[25] Morozov A A and Sushkova O S 2018 A Virtual Machine for Low-Level Video Processing in Actor Prolog *Journal of Physics: Conference Series* **1096** 012044 DOI: 10.1088/1742-6596/1096/1/012044

[26] Morozov A A and Sushkova O S 2018 *Advances in Artificial Intelligence – IBERAMIA* (Cham: Springer International Publishing) 29-41

[27] Morozov A A and Sushkova O S 2019 *The intelligent visual surveillance logic programming* URL: http://www.fullvision.ru

[28] Morozov A A, Sushkova O S, Petrova N G, Khokhlova M N and Migniot C 2018 *Radioelektronika. Nanosistemy. Informacionnye Tehnologii* **10** 101-116

[29] Morozov A A, Sushkova O S, Polupanov A F, Antsiperov V E, Mansurov G K, Paprotskiy S K, Yanushko A V, Petrova N G and Bugaev A S 2018 *Radioelektronika. Nanosistemy. Informacionnye Tehnologii* **10** 311-322

[30] LeCun Y, Bottou L, Bengio Y and Haffner P 1998 *Proceedings of the IEEE* **86** 2278-2324

[31] Krizhevsky A, Sutskever I and Hinton G E 2012 *Advances in Neural Information Processing Systems* **25** 1097-1105

[32] He K, Zhang X, Ren S and Sun J 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* **1** 770-778

[33] Redmon J and Farhadi A 2016 *CoRR* URL: http://arxiv.org/abs/1612.08242

[34] Barmpoutis A 2013 *IEEE Transactions on Cybernetics* **43** 1347-1356

Acknowledgments
Authors are grateful to Renata A. Tolmacheva for the help in the preparation of terahertz/3D video samples and Angelos Barmpoutis for his J4K library [34] which was used for the data collection. Authors thank Dmitry M. Murashov, Feodor D. Murashov, Viacheslav E. Antsiperov, Gennady K. Mansurov, Stanislav K. Paprotskiy, Andrei P. Gorchakov, Alexander V. Yanushko, Nadezda G. Petrova, and Alexander S. Bugaev for cooperation. We are grateful to the Astrohn Technology Ltd and OOO ASoft who provided us with the THERZ-7A terahertz scanning device. The work was carried out within the framework of the state task. This research was partially supported by the Russian Foundation for Basic Research (project number 16-29-09626-ofi-m).