IJMTST is UGC Approved Journal with Journal Id: 43137

International Journal for Modern Trends in Science and Technology, 8(S08): 92-99, 2022

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST08S0817
Available online at: https://www.ijmtst.com/vol8si08.html

Check for updates

Object Localization Based on a Fast Recurrent Regional Based Network (OL-FR-RBN)

Sakthi Priyanka.V | Dr.V.Radha

Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore.

To Cite this Article
Sakthi Priyanka.V and Dr.V.Radha. Object Localization Based on a Fast Recurrent Regional Based Network (OL-FR-RBN). International Journal for Modern Trends in Science and Technology 2022, 8(S08), pp. 92-99. https://doi.org/10.46501/IJMTST08S0817

Article Info
Received: 26 May 2022; Accepted: 24 June 2022; Published: 28 June 2022.

ABSTRACT

Image recognition includes the categories of object localization, object identification, and image categorization. Despite the fact that these terms are sometimes used interchangeably, object detection and object localization are not the same. Similar to this, categorizing photographs and pinpointing their locations are two separate processes that shouldn't be mixed up. It's important to evaluate how well these models work with data points that haven't been seen before while creating machine learning models for use in the real world. To assess how well different models perform when applied to actual data points, we use an assessment metric. The most popular way for gauging the success of an image localization operation is IoU. "Intersection over Union" is referred to as IoU. In this work, object localization was carried out utilizing a Fast Recurrent Reginal Based Network (OL-FR-RPN). The accuracy and recall of each of the best-matching bounding boxes for the recognised objects in the image are used to measure how good an proposedOL-FR-RPN object recognition model is.

KEYWORDS: object detection, classification, localisation, Neural network, RCNN, F-CNN, accuracy, precision, recall, f-measure

1. INTRODUCTION

Object recognition," a catch-all phrase for a number of interconnected computer vision-related activities, is the process of identifying objects in digital pictures. The practice of classifying items based on how they seem is known as image classification. Locating one or more items in an image and then indicating their size by drawing a box around their perimeters is the process of object localization. The process of object detection pinpoints the sort and location of an object within a picture. The object is additionally grouped. Real-time video analysis has grown and changed in a surprising way during the last ten years. The ultimate goal of video analytics is to quickly or automatically spot potentially dangerous situations. Video surveillance is one of the most popular fields of study nowadays. This type of research involves watching people's behavior and classifying it as typical (normal), atypical (abnormal), or suspicious. The main goal is to spot unusual events in movies using a monitoring system that, depending on the situation, may be entirely human, slightly automated, or completely automated. The operator is the only source of control for surveillance systems that need human input. Analyzing behaviour and differentiating between normal and abnormal behaviour involves physical labour in addition to other things. Compared to semi-automatic systems, which depend on human decision-making, fully autonomous video surveillance systems are more sophisticated and smarter. Semi-automatic video surveillance systems, on the other hand, have the ability to make choices without human input. Corporations, governments, and both the public and private sectors are investing more money to improve the security of structures such as workplaces, shopping centres, residences, and other types of infrastructure. According to recent market research, this trend will intensify over the next few years as the autonomous security market grows. Being alert for odd or questionable behaviour that could cause individuals problems is more important than ever. At the moment, terrorist attacks are growing exponentially.

Anomalous events are happenings that defy expectations and the established order. Examples include dropped objects, objects in restricted areas, objects in unusual locations, unusual motion patterns like moving in the wrong direction, illegal traffic turns, human violence or fighting, left bags, and different movements of objects during army training like all soldiers walking but some crawling or walking in different directions. Unexpected objects, irrational object entry or exit, and motion patterns such moving in the wrong direction. It's possible for something to be viewed as unusual in one situation yet normal in another. This is dependent on the current situation. According to a new study, systems that use deep learning to detect and identify strange events outperform those that rely on traditional techniques. (Needs citation) (Needs citation) The progress made by many researchers and the ideas found in the literature to address these issues were validated in this study, which looked at a number of recent research publications on identifying abnormal events. Established ideas won't yield the results that upcoming researchers had hoped for if they are not rigorously examined. If researchers do not conduct a thorough analysis of the current success indicators in their field, they risk missing opportunities for potential future breakthroughs that could result in brilliance.

It is programmed to function toward a particular goal after obtaining data from various monitoring systems. According to experts, watching and recording people's behaviour is the most common use of video surveillance. Establishing a monitoring programme is crucial in order to locate and look at specific incidents before passing judgement. Finding movie characters primarily involves detection, with categorization coming in a distant second. Depending on the surroundings, object detection may be static or dynamic. Object-based classification, object-based classification based on motion, object-based classification based on texture, and object-based classification based on hue are the four main ways to categorise objects in addition to object detection. The most common approach to classifying items is object-based classification. With features like action detection, silhouette extraction, and three-dimensional sensing, to name a few, developers may now create apps for video surveillance systems. Intelligent video surveillance systems are getting smarter thanks to classifiers, machine learning algorithms, and deep learning algorithms, which are all ways to learn. All potential uses for video surveillance are listed in detail in Table 10.4. You must use artificial intelligence, deep learning, and correctly sized items in conjunction with the programme to produce the required results. Before it can be made accessible to the general public, it still needs to go through a few development stages and address a few problems. According to a study on video surveillance and the Internet of Things, the three most crucial problems that must be solved are energy loss, informal latency, and the quality of multimodal data. Some of the issues that have been mentioned include the following: A prototype will need a lot of processing power to handle high-quality multimedia data. As a result of numerous power-saving strategies, it's probable that video quality may decrease and data processing time will increase. These are both undesired results. Therefore, factors like power consumption and image quality must be taken into account while creating a video surveillance Internet of Things system. Given that delayed information is completely useless in an emergency, it is equally crucial to take into account how long it takes to receive the information.

2. REVIEW OF LITERATURE

An adaptive Gaussian mixture model was proposed by Stauffer and Grimson that adapts to changes in dynamic surroundings brought on by changes in lighting, unforeseen events, and other causes. In particular, this model can react to changes in dynamic scenes with accuracy. The results of their investigations show that this model is sensitive to changes in dynamic situations brought on by differences in illumination, outside events, and other factors. Chen and co. The purpose of doing this was to enhance the capability of object tracking and identification. Foreground elements were located and tracked using this model. using the mean-shift technique, which is detailed further down this page. The backdrop photos are segmented using this technique. A system designed and tested by Yilmaz [44] is able to recognise actual human behaviours shown in uncalibrated video clips. A technique for identifying activities based on image edges and highlights was developed by Yamato et al. They called their approach "Activity Recognition." The idea of temporary layouts was developed by Bobick et al.with the intention of recognising activities. In this step, the photos' human-shaped covering is removed, and the variations around the margins are documented. It might be possible to develop an activity model utilising space-time forms obtained from outline data, according to Cover and his colleagues . The Poisson condition is used to find intriguing qualities, such as the salience of the neighbourhood, the direction, the form structure, and the activity characteristics, which are then further studied, based on the outline data and the foundation deduction.

A new, important, and uncomplicated technique for modelling action recognition from profundity successions was created by Omar Oreifej et al. (2013) Profundity successions used this description as a call to action. In addition to a histogram of normal direction in the 3D domain of depth, time, and space, this descriptor collects movement and computation signals in the 4D domain of depth, time, and space. The progression of time and space is also taken into account in this depiction. Darrell and Pentland thought they could figure out how motions are represented by looking at a variety of models from different sources. The many kinds and combinations of connection scores between view models and picture outlines are used to create a coordinated motion format using the DTW computation. For coupling and training, hidden Markov models (HMMs) are used to simulate how different processes interact, as suggested by Brand et al.. In a visual test to identify two-handed movements, hidden Markov models (HMMs) outperformed conventional Markov models. Oliver et al. made the first substantial contribution to prototype communications using the coupled-HMM (CHMM) approach. The results of running two HMM models for two different managers, as well as other hidden state possibilities, will be covered in this section. A straightforward method for differentiating between human activities and 3D joint locations was devised by Yang et al. (2012) . Real-time data on the joints of the human body may now be obtained thanks to the availability of RGBD sensors and the related software development kit (SDK). In situations where intriguing behaviours appeared frequently and randomly among dull activities while those activities were occurring, Dawar et al methods for recognising human activities using depth cameras. The technique organises the activities of interest that have been found after the process of discovering them and deciding which are exciting and which are not. In order to generate an interest point indicator based on global data, such as how pixels are connected throughout an entire video sequence, Wong et al. used non-negative framework factorization in general video grouping. This was done in order to develop a worldwide interest point indicator based on data.

Object Localization Based On Fast Recurrent Reginal Based Network(OL-FR-RPN)

Suggesting candidate regions or bounding boxes of prospective items in a picture is a technique known as "selective search" in computer vision. However, a variety of area proposal strategies can be used because the design is flexible.The feature extractor used in the model was AlexNet deep CNN, which took first place in the 2012 ILSVRC image classification competition. The CNN generated a 4,096-element vector that describes what is in the image using a linear support vector machine (SVM), which is used for classification. One SVM is specifically trained for each recognised class.

On average, using CNNs to identify an object's position and kind is a simple and easy operation. The CNN-based feature extraction phase is required for each of the potential regions produced by the region proposal algorithm, lengthening the procedure. One of the method's drawbacks is this. The model could process about 2,000 possible regions per image during testing, according to the research.The model needs a picture and a list of probable regions as inputs. A deep convolutional neural network is then used on both of these. To extract features, a CNN that has previously been trained is employed, such as a VGG-16. RoI Pooling, or simply Region of Interest Pooling, is the layer that sits at the very top of the deep CNN. The features specific to each potential input location must be extracted by this layer.The model splits into two outputs after a fully connected layer analyses the CNN's output: one for class prediction via a softmax layer and another for the bounding box via a linear output. Then, for each area of interest in the inspected image, this process is repeated several times.

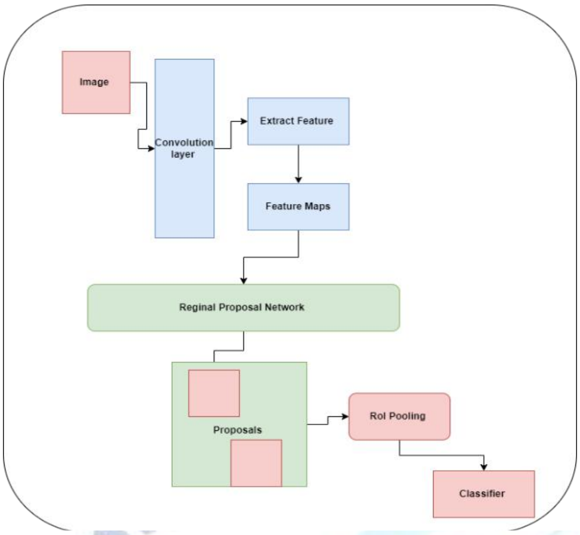
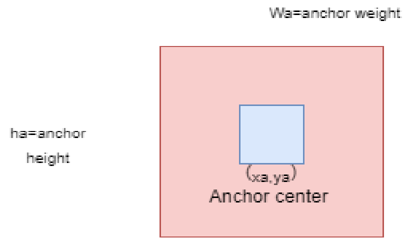


FIGURE 1:ARCHITECTURE OF OL-FR-RPN

The architecture, which we will refer to as an RPN (Area Proposal Network), was created to make new region ideas and improve those that already exist as part of the training process. The proposed model's feature was extracted using Equation 1,

$$n \frac{((V1 - F1 + 2pd)}{p + 1}^{*}$$

where V1 is the input volume's size, F1 is the neurons' receptive field size, p is their applied stride, Pd is the amount of zero padding utilised on the border, and K is their depth. To create a single model design, these results are then combined with those of a Fast R-CNN model. The number of proposed areas has decreased as a result of these improvements, and the model's test-time operation has been sped up to operate more quickly. Cutting-edge performance is the result of this.Then, one of the objects, such as a cat, dog, human, etc., is allocated one of these candidate boxes, which may or may not contain our stuff. The word "classify" has this meaning. In order to better fit the actual object, the box's shape is changed at this time. "Bounding box" is the name for this type of regression.Making candidate boxes is the initial step, which RPN is currently in charge of finishing. Older iterations of object detectors generated offline proposals using standard computer vision algorithms. Selective search is a name for one of them. These systems have two drawbacks: the cost of computation and the incapability of offline calculation.RPN came to the rescue by doing this in very little time and also it can be merged to any object detection network which makes it useful for end-to-end training. Just like how our CNNs learn classification from feature maps, it also learns the proposals from feature maps. In OL-FR-RPN Every point in the feature map generated by the backbone network is an anchor point. We need to generate anchor boxes for every anchor point. We generate candidate boxes using two parameters — scales and aspect ratios. The boxes need to be at image dimensions, whereas the feature map is reduced depending on the backbone.



The fact that RPN can be integrated into any object detection network makes it appropriate for end-to-end training, and it saved the day by finishing this task swiftly. Similar to how our OL-FR-RPN learns how to classify objects using feature maps, it picks up knowledge of the concepts. At each of the locations that make up the feature map of the backbone network, there are anchor points in the OL-FR-RPN. For each and every anchor point, we must build an anchor box. The two factors used to construct possible boxes are size and aspect ratio. Even though the boxes' dimensions must match those of the image, the size of the boxes on the feature map may be less depending on the backbone. The output of a pre-trained deep CNN, such as VGG-16, and the feature map are passed via a small network by the RPN in order to operate. With this method, a list of suggested regions is produced along with a classification prediction for each. The bounding boxes for proposed regions are created using anchor boxes, which are pre-made templates created to simplify and enhance the process of proposing new areas. An object either exists or does not exist in the designated area, depending on whether the class prediction is true or false. This is referred to as the area's "objectness."

Confidence scores are calculated by multiplying the probability of an object by the amount due.

confident scores(CS)= probability of an object * IOU

The abbreviation IOU stands for intersection over union. The following is a formula for calculating the overlapping area of the ground truth and predicted areas in relation to the total area:

IOU = overlapped area / unionised area in this calculation.

A procedure of alternating training is used where both sub-networks are trained at the same time, although interleaved. This allows the parameters in the feature detector deep CNN to be tailored or fine-tuned for both tasks at the same time.

3. PERFORMANCE RESULT ANALYSIS

The effectiveness of an image classification model is assessed using the mean classification error over all projected class labels. The performance metric used for single-object localization is the separation between the anticipated and expected bounding boxes for the predicted class. On the other hand, the precision and recall for each of the best matching bounding boxes for the known objects in the image are examined in order to assess the performance of an object recognition model. A comparison of the proposed OL-FR-accuracy RNP's to that of other neural networks is shown in Table 1. The proposed model astonishingly exceeds the accuracy criteria by 88 per cent. A breakdown of its accuracy is shown in Figure 2.

Table 1 Accuracy comparison

Methods	Accuracy (%)
CNN	85.2
R-CNN	60.9
FR-CNN	87.8
Proposed OL-FR-RPN	88

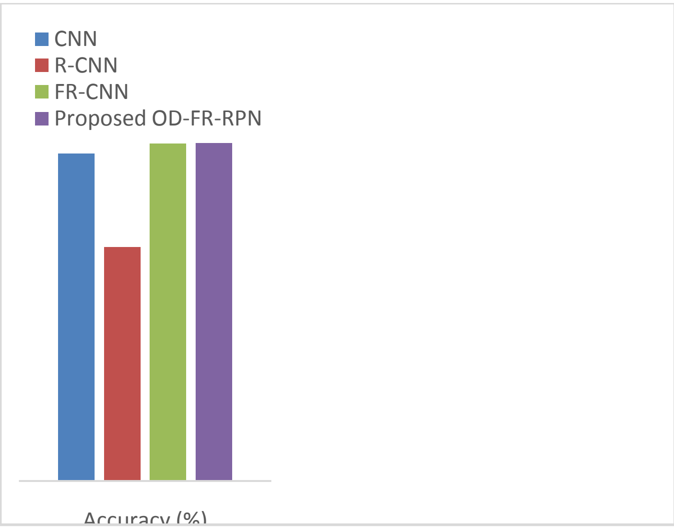


Figure 2 Accuracy Comparison Results

Table 2 describes the Precision Analysis of the proposed OD-FR-RNP with another related neural network. the proposed model achieves a high Precision rate of 89%. Figure 2 Analysis of the Precision

Table 3 Precision comparison results of Adaboost, SVM, deep forest and modified deep forest

Methods	**Precision**
CNN	86
R-CNN	61.9
FR-CNN	88
Proposed OD-FR-RPN	89

Table 3 describes the Recal Analysis of the proposed OL-FR-RNP with other related neural network . the proposed model achives the high Precision rate on on 89%. The figure 3 Analysis the Recal

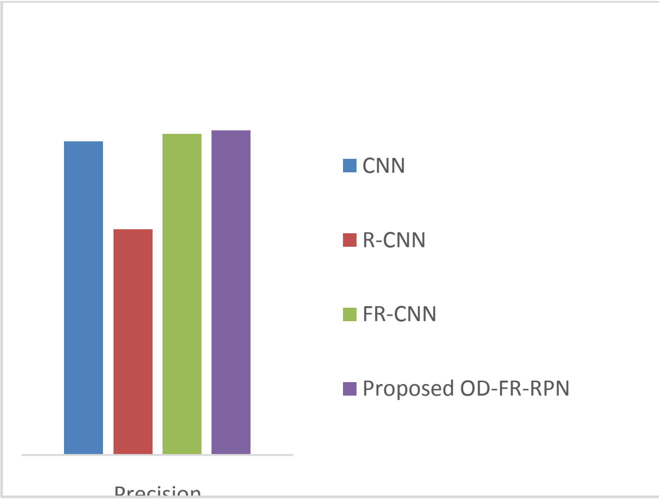


Table 4 Recall comparison results of Adaboost, SVM, deep forest and modified deep forest

Methods	**Recall**
CNN	86.7
R-CNN	77.4
FR-CNN	88
Proposed OL-FR-RPN	89

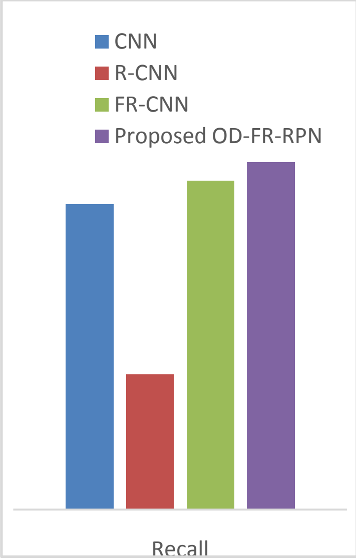


Figure 3 F-measure comparison Results

table 3 describes the F-Measure Analysis of the proposed OD-FR-RNP with other related neural network . the proposed model achives the high F-measure rate on on 89%. The figure 3 Analysis the F-Measure

Table 4. F-measure comparison results of Adaboost, SVM, deep forest and modified deep forest

Methods	**F-Measure**
CNN	87.1
R-CNN	88.79
FR-CNN	88.3
Proposed OL-FR-RPN	89

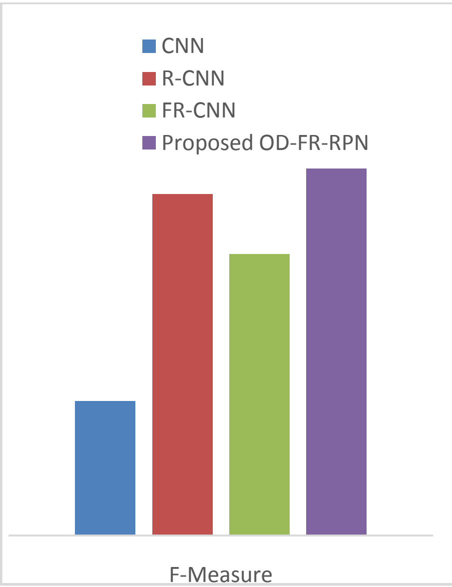


Figure 5 F-measure

4. Conclusion

Public video monitoring is widely used in many countries both by the private and public sectors. It is an essential tool for monitoring people's movements and thwarting crime and terrorism. Both the public and private sectors are affected by this. The need for effective surveillance in public spaces like airports, train stations, shopping malls, crowded sports arenas, and military installations has grown along with the number of people concerned about global security. The adoption of video surveillance by sophisticated healthcare organisations, such as senior living homes, to monitor everyday activities and detect falls, is expected to continue this trend for the foreseeable future. Deep learning has become known as a field of study that may be beneficial recently. According to a new study, systems that use deep learning to detect and identify strange events outperform those that rely on traditional techniques. (Needs citation) (Needs citation) The progress made by many researchers and the ideas found in the literature to address these issues were validated in this study, which looked at a number of recent research publications on identifying abnormal events. Established ideas won't yield the results that upcoming researchers had hoped for if they are not rigorously examined. In this work, object localization was carried out using A Fast Recurrent Reginal Based Network (OL-FR-RPN). The accuracy and recall of each of the best-matching bounding boxes for the recognised objects in the image are used to gauge how good an object recognition model is.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

References

[1] J. Candamo, M. Shreve, D.B. Goldgof, D.B. Sapper, R. Kasturi, Understanding transit scenes: a survey on human behavior-recognition algorithms, IEEE Trans. Intell. Transp. Syst. 11 (1) (2010) 206224.

[2] M. Casares, S. Velipasalar, Adaptive methodologies for energy-efficient object detection and tracking with battery-powered embedded smart cameras, IEEE Trans. Circuits Syst. Video Technol. 21 (10) (2011) 14381452.

[3] A. Bader, H. Ghazzai, A. Kadri, M.S. Alouini, Front-end intelligence for large-scale application-oriented Internet-of-Things, IEEE Access 4 (2016) 32573272. REFERENCES 153

[4] C. Perera, C.H. Liu, S. Jayawardena, M. Chen, A survey on Internet of Things from industrial market perspective, IEEE Access 2 (2014) 16601679.

[5] S.J. Desai, M. Shoaib, A. Raychowdhury, An ultra-low power, "Always-On" camera front-end for posture detection in body worn cameras using restricted boltzman machines, IEEE Trans. Multi-Scale Comput. Syst. 1 (4) (2015) 187194.

[6] M. Mohammadi, A. Al-Fuqaha, S. Sorour, M. Guizani, Deep learning for IoT big data and streaming analysis: a survey, IEEE COM S&T J. (2017) 134.

[7] Y. Sun, H. Song, A.J. Jara, R. Bie, Internet of Things and big data analytics for smart and connected communities, IEEE Access 4 (2016) 766773.

[8] A.K. Singh, S. Rajoriya, S. Nikhil, T.K. Jain, Design constraint in single-hop and multi-hop wireless sensor network using different network model architecture, in: 2nd International Conference on Computing, Communication & Automation, Noida, 2015, pp. 436441

[9] A. Mtibaa, A. Emam, K.A. Harras, On practical multihop wireless communication: insights, limitations, and solutions, in: 2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Rome, 2017, pp. 18.

[10] M.J. Wills, Decision through data: analytics in healthcare, J. Healthc. Manag. 57 (4) (2014) 254262.

[11] C.W. Huang, S. Narayanan, Comparison of feature-level and kernel-level data fusion methods in multisensory fall detection, in: 2016 IEEE 18th International Workshop on MultimediaSignal Processing (MMSP), Montreal, QC, 2016, pp. 16.

[12] Z. Jixian, Multi-source remote sensing data fusion: status and trends, Int. J. Image Data Fusion 1 (1) (2010) 524.

[13] B.N. Silva, M. Khan, K. Han, Internet of Things: a comprehensive review of enabling technologies, architecture, and challenges, IETE Tech. Rev. (2017). Available from: https://doi.org/10.1080/02564602.2016.1276416.

[14] G.A. Akpakwu, B.J. Silva, G.P. Hancke, A.M. Abu-Mahfouz, A survey on 5G networks for the Internet of Things: communication technologies and challenges, IEEE Access 6 (2018) 36193647.

[15] S. Amendola, R. Lodato, S. Manzari, C. Occhiuzzi, G. Marrocco, RFID technology for IoT-based personal healthcare in smart spaces, IEEE Internet Things J. 1 (2) (2014) 144152.

[16] E´. Morin, M. Maman, R. Guizzetti, A. Duda, Comparison of the device lifetime in wireless networks for the Internet of Things, IEEE Access 5 (2017) 70977114.

[17] Y. Zhang, L. Sun, H. Song, X. Cao, Ubiquitous WSN for healthcare: recent advances and future prospects, IEEE Internet Things J. 1 (4) (2014) 311318.

[18] L. Catarinucci, D. de Donno, L. Mainetti, L. Palano, L. Patrono, M.L. Stefanizzi, et al., An IoT-aware architecture for smart healthcare systems, IEEE Internet Things J. 2 (6) (2015) 515526.

[19] Z. Yan, X. Yu, W. Ding, Context-aware verifiable cloud computing, IEEE Access 5 (2017) 22112227.

[20] M. Yousif, Cloudy, foggy and edgy, IEEE Cloud Comput. 4 (2) (2017) 45.

[21] T. Mekonnen, P. Porambage, E. Harjula, M. Ylianttila, Energy consumption analysis of high quality multi-tier wireless multimedia sensor network, IEEE Access 5 (2017) 1584815858.

[22] T. Mekonnen, E. Harjula, T. Koskela, M. Ylianttila, sleepyCAM: power management mechanism for wireless video-surveillance cameras, in: 2017 IEEE International Conference on Communications Workshops (ICC Workshops), Paris, 2017, pp. 9196.

[23] V. Jelici ˇ c, M. Magno, D. Brunelli, V. Bilas, L. Benini, An energy efficient multimodal wireless video ´ sensor network with eZ430RF2500 modules, in: 5th International Conference on Pervasive Computing and Applications, Maribor, 2010, pp. 161166. 152 CHAPTER 10 REVIEW OF INTELLECTUAL VIDEO SURVEILLANCE

[24] R. Bhatt, R. Datta, A two-tier strategy for priority based critical event surveillance with wireless multimedia sensors, Wirel. Netw. 22 (1) (2016) 267284.

[25] C. Long, Y. Cao, T. Jiang, Q. Zhang, Edge computing framework for cooperative video processing in multimedia IoT systems, IEEE Trans. Multimedia 20 (5) (2018) 11261139. Available from: https://doi. org/10.1109/TMM.2017.2764330.

[26] S.K. Sharma, X. Wang, Live data analytics with collaborative edge and cloud processing in wireless IoT networks, IEEE Access 5 (2017) 46214635.

[27] T.W. Martin, K. Chang, A data fusion formulation for decentralized estimation predictions under communications uncertainty, in: 2006 9th International Conference on Information Fusion, Florence, 2006, pp. 17.

[28] L. Sun, Y. Li, R.A. Memon, An open IoT framework based on microservices architecture, China Commun. 14 (2) (2017) 154162.

[29] J.W. Elmenreich, R. Leidenfrost, Fusion of heterogeneous sensors data, in: 2008 International Workshop on Intelligent Solutions in Embedded Systems, Regensburg, 2008, pp. 110.

[30] F. Alam, R. Mehmood, I. Katib, N.N. Albogami, A. Albeshri, Data fusion and IoT for smart ubiquitous environments: a survey, IEEE Access 5 (2017) 95339554.

[31] U. Rashid, H.D. Tuan, P. Apkarian, H.H. Kha, Multisensor data fusion in nonlinear Bayesian filtering, in: 2012 Fouth International Conference of Communication Electronics, 2012, pp. 351354.

[32] M. Marjani, F. Nasaruddin, A. Gani, A. Karim, I.A.T. Hashem, A. Siddiqa, et al., Big IoT data analytics: architecture, opportunities, and open research challenges, IEEE Access 5 (2017) 52475261.

[33] D. He, S. Zeadally, An analysis of RFID authentication schemes for Internet of Things in healthcare environment using elliptic curve cryptography, IEEE Internet Things J. 2 (1) (2015) 7283.

[34] T. Mekonnen, E. Harjula, A. Heikkinen, T. Koskela, M. Ylianttila, Energy efficient event driven video streaming surveillance using sleepyCAM, in: Proceedings of the 2017 IEEE International Conference in Computer and Information Technology, 2017, pp. 107113.

[35] Y.C. Wang, Y.F. Chen, Y.C. Tseng, Using rotatable and directional (R&D) sensors to achieve temporal coverage of objects and its surveillance application, IEEE Trans. Mobile Comput. 11 (8) (2012) 13581371.

[36] G. Zhao, H. Ma, Y. Sun, H. Luo, X. Mao, Enhanced surveillance platform with low-power wireless audio sensor networks, in: 2011 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, Lucca, 2011, pp. 19.

[37] S. Saraceni, A. Claudi, A.F. Dragoni, An active monitoring system for real-time face-tracking based on mobile sensors, in: Proceedings ELMAR-2012, Zadar, 2012, pp. 5356.

[38] A.E. Susu, Low-cost distributed video surveillance with discarded mobile phones, in: 2017 21st International Conference on Control Systems and Computer Science (CSCS), Bucharest, 2017, pp. 279286.

[39] B. Mccluskey, Wear it well, Eng. Technol. 12 (1) (2017) 3235.

[40] S.N. Jyothi, K.V. Vardhan, Design and implementation of real time security surveillance system using IoT, in: 2016 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, 2016, pp.15.

[41] Chen Y-T, Chu-Song C, Chun-Rong H, Yi-Ping H: Efficient hierarchical method for background subtraction. Pattern Recognit. 2007, 40: 2706-2715. 10.1016/j.patcog.2006.11.023

[42] Quan S, Zhixing T, Songchen H: Hierarchical CodeBook for background subtraction in MRF. Infrared Phys. Technol. 2013, 61: 259-264.

[43] Kim K, Chalidabhongse TH, Harwood D, Davis L: Real-time foreground–background segmentation using codebook model. Real-time Imaging 2005, 11(3):172-185. 10.1016/j.rti.2004.12.004

[44] Alper Yilmaz and Mubarak Shah, "Recognizing Human Actions in Videos Acquired by Uncalibrated Moving Cameras". Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, 17-21 Oct. 2005

[45] Yamato, J., Ohya, J., & Ishii, K. (1992, June). Recognizing human action in time-sequential images using hidden Markov model. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR '92 (pp. 379-385). IEEE. doi:10.1109/CVPR.1992.223161

[46] Bobick, A. F., & Davis, J. W. (2001). The recognition of human movement using temporal templates. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(3), 257–267.

[47] Blank, M., ... (2005). Actions as space-time shapes. Proceedings of the Tenth IEEE International Conference on Computer Vision ICCV '05 (Vol. 2). IEEE. doi:10.1109/ICCV.2005.28.

[48] Omar Oreifej, Zicheng Liu; HON4D: Histogram of Oriented 4D Normals for Activity Recognition from Depth Sequences, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 716-723.

[49] T. Darrell and A. Pentland, "Space-time gestures," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 335–340, New York, NY, USA, 1993.

[50] Matthew Brand, Nuria Oliver, and Alex Pentland Coupled hidden Markov models for complex action recognition. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 7-19 June 1997E. H. Miller, "A note on reflector arrays (Periodical style—Accepted for publication)," IEEE Trans. Antennas Propagat., to be published.