Metadata of the chapter that will be visualized in SpringerLink

Book Title	Intelligent Data Engineering and Automated Learning – IDEAL 2020	
Series Title		
Chapter Title	Review of Trends in Automatic Human Activity Recognition Using Synthetic Audio-Visual Data	
Copyright Year	2020	
Copyright HolderName	Springer Nature Switzerland AG	
Author	Family Name	**Jesus**
	Particle	
	Given Name	**Tiago**
	Prefix	
	Suffix	
	Role	
	Division	Centre Algoritmi
	Organization	University of Minho
	Address	4710-057, Braga, Portugal
	Division	
	Organization	Bosch Car Multimedia
	Address	4705-820, Braga, Portugal
	Email	
	ORCID	http://orcid.org/0000-0003-1437-5439
Author	Family Name	**Duarte**
	Particle	
	Given Name	**Júlio**
	Prefix	
	Suffix	
	Role	
	Division	Centre Algoritmi
	Organization	University of Minho
	Address	4710-057, Braga, Portugal
	Division	
	Organization	Bosch Car Multimedia
	Address	4705-820, Braga, Portugal
	Email	
	ORCID	http://orcid.org/0000-0002-5458-3390
Author	Family Name	**Ferreira**
	Particle	
	Given Name	**Diana**
	Prefix	
	Suffix	
	Role	
	Division	Centre Algoritmi

	Organization	University of Minho
	Address	4710-057, Braga, Portugal
	Division	
	Organization	Bosch Car Multimedia
	Address	4705-820, Braga, Portugal
	Email	
	ORCID	http://orcid.org/0000-0003-2326-2153
Author	Family Name	**Durães**
	Particle	
	Given Name	**Dalila**
	Prefix	
	Suffix	
	Role	
	Division	Centre Algoritmi
	Organization	University of Minho
	Address	4710-057, Braga, Portugal
	Division	
	Organization	Bosch Car Multimedia
	Address	4705-820, Braga, Portugal
	Email	
	ORCID	http://orcid.org/0000-0002-8313-7023
Author	Family Name	**Marcondes**
	Particle	
	Given Name	**Francisco**
	Prefix	
	Suffix	
	Role	
	Division	Centre Algoritmi
	Organization	University of Minho
	Address	4710-057, Braga, Portugal
	Division	
	Organization	Bosch Car Multimedia
	Address	4705-820, Braga, Portugal
	Email	
	ORCID	http://orcid.org/0000-0002-2221-2261
Author	Family Name	**Santos**
	Particle	
	Given Name	**Flávio**
	Prefix	
	Suffix	
	Role	
	Division	Centre Algoritmi
	Organization	University of Minho
	Address	4710-057, Braga, Portugal
	Division	

	Organization	Bosch Car Multimedia
	Address	4705-820, Braga, Portugal
	Email	
	ORCID	http://orcid.org/0000-0003-2378-5376
Author	Family Name	**Gomes**
	Particle	
	Given Name	**Marco**
	Prefix	
	Suffix	
	Role	
	Division	Centre Algoritmi
	Organization	University of Minho
	Address	4710-057, Braga, Portugal
	Division	
	Organization	Bosch Car Multimedia
	Address	4705-820, Braga, Portugal
	Email	
	ORCID	http://orcid.org/0000-0001-6370-9955
Author	Family Name	**Novais**
	Particle	
	Given Name	**Paulo**
	Prefix	
	Suffix	
	Role	
	Division	Centre Algoritmi
	Organization	University of Minho
	Address	4710-057, Braga, Portugal
	Division	
	Organization	Bosch Car Multimedia
	Address	4705-820, Braga, Portugal
	Email	
	ORCID	http://orcid.org/0000-0002-3549-0754
Author	Family Name	**Gonçalves**
	Particle	
	Given Name	**Filipe**
	Prefix	
	Suffix	
	Role	
	Division	
	Organization	Bosch Car Multimedia
	Address	4705-820, Braga, Portugal
	Email	
	ORCID	http://orcid.org/0000-0002-8769-4257
Author	Family Name	**Fonseca**
	Particle	

	Given Name	**Joaquim**
	Prefix	
	Suffix	
	Role	
	Division	
	Organization	Bosch Car Multimedia
	Address	4705-820, Braga, Portugal
	Email	
	ORCID	http://orcid.org/0000-0002-2056-1206
Corresponding Author	Family Name	**Lori**
	Particle	
	Given Name	**Nicolas**
	Prefix	
	Suffix	
	Role	
	Division	Centre Algoritmi
	Organization	University of Minho
	Address	4710-057, Braga, Portugal
	Division	
	Organization	Bosch Car Multimedia
	Address	4705-820, Braga, Portugal
	Email	nicolas.lori@algoritmi.uminho.pt
	ORCID	http://orcid.org/0000-0002-5895-0880
Author	Family Name	**Abelha**
	Particle	
	Given Name	**António**
	Prefix	
	Suffix	
	Role	
	Division	Centre Algoritmi
	Organization	University of Minho
	Address	4710-057, Braga, Portugal
	Division	
	Organization	Bosch Car Multimedia
	Address	4705-820, Braga, Portugal
	Email	abelha@di.uminho.pt
	ORCID	http://orcid.org/0000-0001-6457-0756
Author	Family Name	**Machado**
	Particle	
	Given Name	**José**
	Prefix	
	Suffix	
	Role	
	Division	Centre Algoritmi
	Organization	University of Minho

Address	4710-057, Braga, Portugal
Division	
Organization	Bosch Car Multimedia
Address	4705-820, Braga, Portugal
Email	jmac@di.uminho.pt
ORCID	http://orcid.org/0000-0003-4121-6169

Abstract	An in-depth study of knowledge and technologies was made related to the various scientific, technical, and industrial domains necessary for the acquisition of skills and capabilities for the design and development of a multisensory fusion system for vehicle cockpits. After an extensive literature review, it was possible to determine the baselines of the solution to be developed and obtain a pipeline prototype.

Keywords	Autonomous car - Multisensory fusion - Audio-visual synthetic data

Review of Trends in Automatic Human Activity Recognition Using Synthetic Audio-Visual Data

Tiago Jesus[1,2], Júlio Duarte[1,2], Diana Ferreira[1,2], Dalila Durães[1,2], Francisco Marcondes[1,2], Flávio Santos[1,2], Marco Gomes[1,2], Paulo Novais[1,2], Filipe Gonçalves[2], Joaquim Fonseca[2], Nicolas Lori[1,2(✉)], António Abelha[1,2], and José Machado[1,2]

[1] Centre Algoritmi, University of Minho, 4710-057 Braga, Portugal
nicolas.lori@algoritmi.uminho.pt,
{abelha,jmac}@di.uminho.pt
[2] Bosch Car Multimedia, 4705-820 Braga, Portugal

AQ1

Abstract. An in-depth study of knowledge and technologies was made related to the various scientific, technical, and industrial domains necessary for the acquisition of skills and capabilities for the design and development of a multisensory fusion system for vehicle cockpits. After an extensive literature review, it was possible to determine the baselines of the solution to be developed and obtain a pipeline prototype.

Keywords: Autonomous car · Multisensory fusion · Audio-visual synthetic data

1 Introduction

For a primary statement, there are very few researchers focusing on in-vehicle action recognition, presumably due to privacy issues/concerns. Therefore, several elements for this approach will need to be built, at least in part, from scratch. This is the key-risk for this publication and the results will reflect this reality.

1.1 Internet of Things (IoT)

In this novel paradigm, embedded sensors and Internet connectivity are installed in smart objects, serving as facilitators to interactions, communication, and integration with the surrounding environment to provide intelligent and useful services [10].

Focusing on the automotive area, a lot of works have been produced on vehicular communications specifically for inter-vehicular, intra-vehicular and vehicle to infrastructures WSNs (Wireless Sensor Networks). The increasing demand for driver safety and assistance in a modern vehicle, has brought attention to the intra-vehicle WSNs.

C. Analide et al. (Eds.): IDEAL 2020, LNCS 12490, pp. 1–12, 2020.
https://doi.org/10.1007/978-3-030-62365-4_53

As more automotive designers implement IoT into their designs, more and smarter cars are manufactured. A recent trend in the automotive industry are smart cars enabled with IoT. Many researchers [11,12,18,27] identify some of the important implications of IoT in transportation:

i. It allows the use of cloud-based intelligent monitoring control system for tracking the location of vehicles in real time;
ii. It allows the communication between equipment through devices attached to vehicles (inter-equipment connection), allowing drivers to avoid delays or accidents;
iii. It can improve passenger comfort and convenience by alerting them about delays via their mobile devices;
iv. It allows a predictive maintenance by providing to vehicles the capacity of transmitting defect-indication data directly to engineers. Predictive maintenance can identify components in need of repair/replacement.

In this sense, important data can be captured from various IoT connections and devices. For example, sensors installed in vehicles offer the ability to track maintenance needs, driver safety, fuel usage and other related metrics in real time. The data collected from these sensors can help companies to optimize performance and can lead to profitable outcomes for themselves through better user experiences. These data can be used to improve the way that they design, upgrade and maintain devices in the field [23].

The ever-increasing number of connected devices will enable a complex atmosphere with billions of sensors and devices connected to the internet, to ultimately gather, analyze and transmit data in real time. Thus, without the data, IoT would not be able to hold the features and functionalities that have brought them incredible benefits, powerful emphasis and world-wide magnitude.

1.2 Big Data

As cities around the world are increasingly digitized, we are on the verge of an era in which IoT will comprise massive amounts of devices capable of sensing, computing, capturing and operating in the real world [25]. Every day, these devices will generate continuous streams of real-time data on critical infrastructure components and services. The magnitude of the daily explosion of high volumes of data has led to the emerging Big Data paradigm (e.g. [1,2,14–16]).

In a data-driven utopia, data would be highly valued and used in an ethical and effective way. In reality, however, data must travel a long way before it reaches its highest purpose, gaining incremental value as it goes. The essence of the data value chain is to provide a framework through which the data lifecycle can be perceived, transforming low-value inputs (raw data) into high-value outputs (actionable information).

As EU Commissioner Kroes said, "Big Data is the new oil", it is a technology and innovation driver that creates value not only for companies but also for citizens and society [13].

Given the continuous technological advances and the focus towards autonomous driving in recent years, autonomous vehicles are expected to be a major source of Big Data. Over the last few years, the degree of car automation has gradually increased to the point of fully autonomous driving vehicles being a reality in the near future [25]. Using vehicles to form ad-hoc vehicle networks (VANETs) or the Internet of Vehicles (IoVs) has now become a reality. The IoV Big Data is a major enabling technology for conceiving revolutionary self-driving vehicles. As a matter of fact, several researchers [5,6,8,24,28] have already dedicated their efforts towards big sensor data systems using vehicles as sensing elements in a large networked system for intelligent transportation and were able to outline the following insight that to make the self-driving come to life, a convergence of Big Data is required, including the data from on-board sensors, e.g., cameras, radar, Lidar, GPS, and information shared from other connected vehicles, e.g.. road condition and traffic information.

Accordingly, Big Data is a powerful technology that requires attention and investigation to make exciting new mobility features possible and to enable unprecedent vehicular phenomena and experiences while offering efficiency benefits and improving safety.

1.3 Audio-Visual Violence Detection

Figure 1 represents "the basic modes of detection of violence". This figure shows a model for automated detection of human-computer behavior, through selective identification of a person. The method is based on a combination of data from audio, video, and other sensors. The video data is a collection of images of a location. The video data is further processed to extract human and human hierarchical resources [9]. People's actions are detected and processed to detect selectable behavior. A configurable behavior rule can be used to select people's behavior and resources. The hybrid human detector algorithm has been used for human detection, which may include one of several machine learning algorithms.

It's possible to extract six segment-level audio features, which can be used at a next stage by a classifier:

- **Time-Domain Features** express abrupt changes during time of the audio signal. This feature can be calculated by further dividing the frames into N sub-windows of fixed duration;
- **Frequency-Domain Features** expresses abrupt changes in amplitude of the audio signal;
- **Cepstrum** it is an approximately continuous signal, owing to a large part, to the smoothing effect of windowing;
- **Time-frequency** changes in frequency along the time;
- **Energy** admissible poses of the microphones considering the sound energy level and the ILD as acoustic features;
- **Biologically or Perceptually Driven** spectral features based on Gammatone Filter Bank.

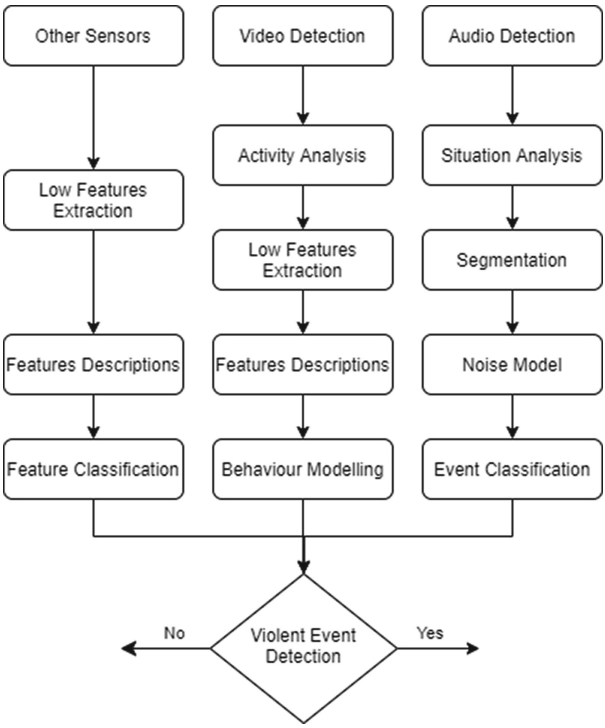

Fig. 1. Basic ways of violence detection.

From the computer vision community point of view, visual tracking is the process of locating, identifying, and determining the dynamic configuration of one or many moving (possibly deformable) objects (or parts of objects) in each frame of one or several cameras. What makes a good visual tracking algorithm is its capacity to handle all the variability in a video sequence caused by the tracked object, the scene and the camera acquiring the scene. Such variability can be caused, for instance, by pose and illumination variations, occlusions, varying and erratic motions [7]. Visual tracking in video sequences raises many problems that can cause a loss of track on the object. Table 1 presents a list of some of the most challenging ones [7].

2 Methods

This section is dedicated to a selection of the datasets publicly available on the web for human activity recognition. There are several datasets, however, we can divide them in several types: visual tracking, motion recognition and action recognition using video and audio.

Table 1. List of some of the most challenging problems while visual tracking objects in video [7].

Problem	Description
Illumination effects	The light scene environment may change due to external conditions (weather, time of day) or internal conditions (lights on or off). Furthermore, depending on the incidence of light, the variations of light can cause object colors to change over time, which can confuse the visual tracking algorithm
Scene clutter	Can happen due to a very textured background or other moving objects in the scene, often similar to the tracked object. This feature can cause some deviations from the visual tracker, resulting in the loss of tracking of the object
Changes in object appearance	Happens because of the projection of 3D movements onto a 2D plane (frames from sequences), the tracked object can have geometric deformations
Abrupt changes in motion	The object velocity can vary with time. This can make the object very hard to track because its movement can become unpredictable and, therefore, the object can be lost
Occlusions	Difficult to handle because some parts of the object can disappear from the scene
Similar appearances	When different objects have similar appearances in the video sequence it may be difficult for the visual tracker to discriminate between these objects

Several public datasets for visual tracking exist such as the datasets: BEHAVE, BoBoT, CAVIAR, Ross, and SPOT. Table 2 describes some of these datasets.

Regarding motion recognition, some examples of dataset are: HDM05, MSR-Daily Activity3D, UTKinect, ActivityNet, MSR-Action3D, RGBD-HuDaAct, CAD-60, MSRC-12, YouTube 8M, and Hollywood 3D. Some of the datasets were analyzed and the information was is summarized in Table 3.

Most of the developed audio scene datasets are not publicly accessible. Table 4 summarizes some of the ones publicly available which are commonly used in audio scene recognition.

There are several dataset repositories for human action. However, it was not possible to find any specific in-vehicle human action dataset. Therefore, a

Table 2. Publicly available datasets used in visual tracking.

Dataset	Description
BoBoT	It presents twelve video sequences in .avi format. All frames have a size of 320×240 pixels and their numbers vary from 305 to 1308. Ground truth commentaries are also given for each sequence. They match the coordinates of the target object's bounding box and its size. This dataset was used in several recent works [7,26]
CAVIAR	The CAVIAR (Context Aware Vision using Image based Active Recognition) project, from MIA Labs was dedicated to the development of algorithms to richly describe and understand video scenes. It contains a lot of information, such as rectangular bounding boxes' locations and sizes, head and feet positions, body direction, etc. The CAVIAR dataset is very popular and used by a lot of computer vision research teams
SPOT	It proposes six very challenging video sequences that were collected from Youtube. It's dedicated to track simultaneously multiple objects, sometimes with similar appearances. However, the movements of the objects in a same sequence are related to each other

Table 3. Publicly available datasets used in motion recognition.

Datasets	References	Classes
MSR-Action3D	[26]	Contains 20 actions: high arm wave, draw x, hammer, hand catch, horizontal arm wave, forward punch, high throw, draw circle, bend, forward kick, side kick, jogging, draw tick, two hand wave, hand clap, tennis serve, pick up, golf swing, throw and side-boxing
UTKinect dataset	[26]	Actions include sit down, stand up, wave, clap hands, walk, throw, push, pick up, carry and pull
ActivityNet	[21]	The dataset divided into three subgroups by application domain such as untrimmed videos classification, trimmed videos classification, and Activity detection on all the untrimmed videos
YouTube 8M	[21]	Provides 7 million videos 12 billion/audio/visual features, 4716 classes, and 3.4 average labels per videos. Each video sequence has 120–500 s in length and more than one thousand views per video
Hollywood 3D	[21]	It consists of 14 activities classes such as no actions, runs, punches, kicks, shots, eat, drive, uses phone, kiss, hug, stands up, sit downs, swims, and dances. There are 650 manually labelled videos in the dataset approximately

dataset must be created from scratch and may require the usage of Synthetic Sensor Data Generation for enabling the proof of concept. A possibility to be explored is to find datasets that produce an approximate result considering the in-vehicle dataset. Nevertheless, this must be explored after a primary dataset is created for comparison.

As pointed out by [22], popular datasets that contain human action have become increasingly detailed over those used for early works. Since then, there have been many advances in how datasets are created. Improvement of quality datasets allows for more complex models, hence the use of challenging datasets allow for the evaluation of the robustness and generalizability of these models. It is crucial, however, to consider that many approaches depend upon inputs that are absent in common/popular datasets, which then require the creation of domain-specific datasets. Meanwhile, some essential characteristics of these popular datasets (that can be found in the literature) must be highlighted. Namely, the main and most relevant can be expressed according to Table 5 [21]:

For the preparation of a human activity behavior dataset from scratch, it is necessary to define a basic structure that respects the most popular characteristics in similar datasets. To set a baseline and what are the proper criteria and workflow for generating a new dataset, some guidelines and recommended best-practices must be followed.

Table 4. Datasets used in Environmental Audio Scene Recognition (EASR) Task [4].

Datasets	Environmental audio scene classes
DARES-G1 (10 contexts)	Basketball game, beach, inside an office facility, inside a bus, grocery shop, inside a car, street, hallways, restaurant, and stadium with track and field events
LITIS-Rouen audio scene dataset (19 contexts)	Busy street, bus, cafe, car, train station hall kid game hall, market, metro-Paris, metro-Rouen, high-speed train, billiard pool hall, quiet street, plane, student hall, restaurant, pedestrian street, shop, train, and tube station
TUT-CASR (18 contexts)	Six high-level classes are: (1) vehicles, (2) outdoors, (3) public/social, (4) offices/meetings/quiet, (5) home, and (6) reverberant places
IEEE/AASP DCASE 2013 Scene dataset (10 contexts)	Bus, busy street, restaurant, office, open-air market, park, quiet street, supermarket, tube (subway train), and tubestation (subway station)
TUT-2016 dataset (15 contexts)	Beach, bus, cafe, car, city center, forest path, grocery store, home, library, metro station, office, residential area, train, tram, and urban park
TUT-2017 dataset (15 contexts)	Bus, cafe/restaurant, car, city center, forest path, grocery store, home, lakeside beach, library, metro station, office, residential area, train (traveling, vehicle), tram (traveling, vehicle), and urban park

3 Results

The state of the art was properly studied, and we concluded that there are five articles that constitute the state-of-the-art and that we now analyse in detail, first the pipeline analysis, and then its data structure.

In [17], a two-stage pipeline was developed to handle pose estimation. The first stage of the proposed cascade model is based in the Faster-RCNN method applied on top of a ResNet-101 CNN to predict the location and scale of boxes likely to contain people. The boxes containing people are then used in the second stage, where the locations of each keypoint for each of the boxes are predicted with a fully convolutional ResNet. A novel keypoint-based Non-Maximum- Suppression (NMS) mechanism is used to avoid duplicate pose detections. Hence building directly on the object keypoint similarity (OKS) metric (called OKS-NMS), instead of a cruder box-level approach called IOU NMS. Finally, a novel keypoint-based confidence score estimator is proposed, which leads to greatly improved average precision compared to using the Faster-RCNN box scores for ranking the final pose proposals.

In [19], two algorithms of the 2016 COCO Keypoints Challenge were evaluated - [17] and [3] - and made four contributions. The first was the taxonomization of the types of error that are typical of the multi-instance pose estimation frameworks. The second contribution was sensitivity analysis of those errors with respect to measures of image complexity. The third was side-by-side comparison of two leading human pose estimation algorithms highlighting key differences in behaviour that are hidden in the average performance numbers. Finally, the last contribution was the assessment of which types of datasets and benchmarks would be most productive in guiding future research.

In [20], a comprehensive survey was presented, as a review paper, of both handcrafted and learning-based action representations, offering comparison, analysis, and discussions on these approaches.

In [29], a weakly-supervised transfer learning method was proposed. It used mixed 2D and 3D labels in a unified deep neural network that presents two-stage cascated structure. The network augmented a 2D pose estimation sub-network by use of a 3D depth regression sub-network.

Table 5. Popular dataset characteristics.

Characteristic	Definition
Classes	The choice of action classes greatly affects the diversity and coverage of the dataset. Although certain datasets construct a rich hierarchy of action classes, in most datasets the following groupings are usually present in some form: Person-Object, Person only and Person-Person
Focus	A majority of the datasets either consider activities performed during daily life or do not have a very specific domain focus

(*continued*)

Table 5. (*continued*)

Characteristic	Definition
Modality	Human action datasets usually preserve the temporal dimension, unlike simpler image classification tasks, although there exist large datasets for image-based activity classification. Singh (Singh, 2019) highlights that the natural representation of datasets is in the form of clips of 2D images, and most datasets use this format extensively. However, the introduction of low-cost 3D sensors such as Microsoft Kinect, has brought great interest in using depth information. A non-visual sensor is used to record an entirely different class of datasets
Data source	The source from which dataset is acquired determines to a large degree how well an algorithm trained on it will perform on unseen data. The datasets consisting mainly of video clips containing one or a pair of actors performing an activity in an indoor lab setting and are called, respectively, recorded or scripted datasets
Annotation method	For generated, recorded, and crowdsourced datasets, the video label is already known at the start of video creation. On the other hand, for annotated datasets, accurate and precise annotation and subsequent verification of labels are essential for supervised learning schemes
Annotation type	The type of annotation determines the degree to which an action is localized in time and space. When temporal localization is not an important concern, as in activity classification problems, the entire sequence or video clip is directly labeled with its corresponding class, this is called a sequence level annotation and is the case for the majority of datasets due to constraints of more complex annotations. For datasets focusing on detecting activities specifically, frame range or action segment annotation specifies an interval related to a particular class
Evaluation	In general, a human action algorithm can have two basic tasks with different evaluation protocols. In action classification task (trimmed activity recognition), the action being performed in a particular clip is identified and thus becomes a multiclass classification problem (with/without null class). In the cases where the sequence level ground truth is available, most datasets use multiclass accuracy as the metric. Although realistic datasets are unbalanced and long-tailed, some alternative metrics such as recall, precision and F score are often used instead

4 Conclusion and Discussion

This work started by introducing the subjects of IoT and Big Data. These are major subjects that need to be addressed when trying to create a pipeline for human activity recognition either inside or outside a vehicle. It then continues by introducing one of the main topics: violence detection. To perform the detection of violence we need to apply human activity recognition algorithms capable of

analysing the faintest signs of violence in a certain situation. Thus, a study was made about the state of the art so as to obtain a better understanding of action recognition.

After studying the literature, we concluded that, although not perfect, a typical pipeline exists to handle automatic human activity recognition. Furthermore, we found several datasets, publicly available whilst searching the literature. To accurately perform action recognition, a two-stage model should be applied as it achieved better results according to the literature.

Finally, by knowing this typical pipeline exists, and having several datasets at hand, we can adapt the pipeline to our specific needs in the future and develop a pipeline prototype to detect situations where violence occurs.

Acknowledgments. This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020. Human and material resources have also been supported by the European Structural and Investment Funds in the FEDER component, through the Operational Competitiveness and Internationalization Programme (COMPETE 2020) [Project number 039334; Funding Reference: POCI-01-0247-FEDER-039334].

References

1. Analide, C., Novais, P., Machado, J., Neves, J.: Quality of knowledge in virtual entities. In: Encyclopedia of Communities of Practice in Information and Knowledge Management, pp. 436–442. IGI Global (2006)
2. Brandão, A., et al.: A benchmarking analysis of open-source business intelligence tools in healthcare environments. Information **7**(4), 57 (2016)
3. Cao, Z., Simon, T., Wei, S.E., Sheikh, Y.: Realtime multi-person 2D pose estimation using part affinity fields. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7291–7299 (2017)
4. Chandrakala, S., Jayalakshmi, S.: Environmental audio scene and sound event recognition for autonomous surveillance: a survey and comparative studies. ACM Comput. Surv. (CSUR) **52**(3), 1–34 (2019)
5. Chaqfeh, M., Lakas, A., Jawhar, I.: A survey on data dissemination in vehicular ad hoc networks. Veh. Commun. **1**(4), 214–225 (2014)
6. Dikaiakos, M.D., Iqbal, S., Nadeem, T., Iftode, L.: VITP: an information transfer protocol for vehicular computing. In: Proceedings of the 2nd ACM International Workshop on Vehicular Ad Hoc Networks, pp. 30–39 (2005)
7. Dubuisson, S., Gonzales, C.: A survey of datasets for visual tracking. Mach. Vis. Appl. **27**(1), 23–52 (2015). https://doi.org/10.1007/s00138-015-0713-y
8. Gerla, M.: Vehicular cloud computing. In: 2012 The 11th Annual Mediterranean Ad hoc Networking Workshop (Med-Hoc-Net), pp. 152–155. IEEE (2012)
9. Gilbert, A., Illingworth, J., Bowden, R.: Action recognition using mined hierarchical compound features. IEEE Trans. Pattern Anal. Mach. Intell. **33**(5), 883–897 (2010)
10. Kim, K.J.: Interacting socially with the internet of things (IoT): effects of source attribution and specialization in human-IoT interaction. J. Comput. Med. Commun. **21**(6), 420–435 (2016)

11. Leng, Y., Zhao, L.: Novel design of intelligent internet-of-vehicles management system based on cloud-computing and internet-of-things. In: Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology, vol. 6, pp. 3190–3193. IEEE (2011)
12. Lumpkins, W.: The internet of things meets cloud computing [standards corner]. IEEE Consum. Electron. Mag. **2**(2), 47–51 (2013)
13. María Cavanillas, J., Curry, E., Wahlster, W.: New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in Europe. Springer Nature (2016)
14. Neto, C., Brito, M., Lopes, V., Peixoto, H., Abelha, A., Machado, J.: Application of data mining for the prediction of mortality and occurrence of complications for gastric cancer patients. Entropy **21**(12), 1163 (2019)
15. Neves, J., Martins, M.R., Vilhena, J., Neves, J., Gomes, S., Abelha, A., Machado, J., Vicente, H.: A soft computing approach to kidney diseases evaluation. J. Med. Syst. **39**(10), 131 (2015)
16. Neves, J., Vicente, H., Esteves, M., Ferraz, F., Abelha, A., Machado, J., Machado, J., Neves, J., Ribeiro, J., Sampaio, L.: A deep-big data approach to health care in the AI age. Mob. Netw. Appl. **23**(4), 1123–1128 (2018)
17. Papandreou, G., et al.: Towards accurate multi-person pose estimation in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4903–4911 (2017)
18. Qin, E., Long, Y., Zhang, C., Huang, L.: Cloud computing and the internet of things: technology innovation in automobile service. In: Yamamoto, S. (ed.) HIMI 2013. LNCS, vol. 8017, pp. 173–180. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39215-3_21
19. Ruggero Ronchi, M., Perona, P.: Benchmarking and error diagnosis in multi-instance pose estimation. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 369–378 (2017)
20. Sargano, A.B., Angelov, P., Habib, Z.: A comprehensive review on handcrafted and learning-based action representation approaches for human activity recognition. Appl. Sci. **7**(1), 110 (2017)
21. Singh, R., Sonawane, A., Srivastava, R.: Recent evolution of modern datasets for human activity recognition: a deep survey. Multimed. Syst. 1–24 (2019)
22. Singh, T., Vishwakarma, D.K.: Video benchmarks of human action datasets: a review. Artif. Intell. Rev. **52**(2), 1107–1154 (2018). https://doi.org/10.1007/s10462-018-9651-1
23. Uden, L., He, W.: How the internet of things can help knowledge management: a case study from the automotive domain. J. Knowl. Manag. **21**, 57–70 (2017)
24. Xu, W., et al.: Internet of vehicles in big data era. IEEE/CAA J. Automatica Sinica **5**(1), 19–35 (2017)
25. Zaslavsky, A., Perera, C., Georgakopoulos, D.: Sensing as a service and big data. arXiv preprint arXiv:1301.0159 (2013)
26. Zhang, J., Li, W., Ogunbona, P.O., Wang, P., Tang, C.: RGB-D-based action recognition datasets: a survey. Pattern Recogn. **60**, 86–105 (2016)
27. Zhang, Y., Chen, B., Lu, X.: Intelligent monitoring system on refrigerator trucks based on the internet of things. In: Sénac, P., Ott, M., Seneviratne, A. (eds.) ICWCA 2011. LNICST, vol. 72, pp. 201–206. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-29157-9_19
28. Zhou, H., et al.: Chaincluster: engineering a cooperative content distribution framework for highway vehicular communications. IEEE Trans. Intell. Transp. Syst. **15**(6), 2644–2657 (2014)

29. Zhou, X., Huang, Q., Sun, X., Xue, X., Wei, Y.: Towards 3D human pose estimation in the wild: a weakly-supervised approach. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 398–407 (2017)

Author Queries

Chapter 53

Query Refs.	Details Required	Author's response
AQ1	This is to inform you that corresponding author has been identified as per the information available in the Copyright form.	