DOI: 10.15514/ISPRAS-2024-36(1)-11

A Systematic Literature Review on Vision-Based Human Event Recognition in Smart Classrooms: Identifying Significant Events and Their Applications

M. L. Córdoba-Tlaxcalteco, ORCID: 0009-0001-8150-8097 <*marcordoba@uv.mx*>
E. Benítez-Guerrero, ORCID: 0000-0001-5844-4198 <*edbenitez@uv.mx*>
Faculty of Statistics and Informatics, University of Veracruz
Xalapa, Veracruz, Mexico

Abstract. The field of vision-based human event recognition in smart environments has emerged as a thriving and successful discipline, with extensive efforts in research and development driving notable progress. This progress has not only yielded valuable insights but also practical applications across various domains. Within this context, human actions, activities, interactions, and behaviors are all considered as events of interest in smart environments. However, when focusing on smart classrooms, a lack of unified consensus on the definition of "human event" poses a significant challenge for educators, researchers, and developers. This lack of agreement hinders their ability to precisely identify and classify specific situations that are relevant to the educational context. To address this challenge, the aim of this paper is to conduct a systematic literature review of significant events, with a particular emphasis on their applications in assistive technology. The review encompasses a comprehensive analysis of 227 published documents spanning from 2012 to 2022. It delves into key algorithms, methodologies, and applications of vision-based event recognition in smart environments. As a primary outcome, the review identifies the most significant events, categorizing them according to single-person behavior, multiple-person interactions, or object-person interactions, examining their practical applications within the educational context. The paper concludes with a discussion on the relevance and practicality of vision-based human event recognition in smart classrooms, especially in the post-COVID era.

Keywords: human event recognition, smart classroom, computer vision, artificial intelligence, educational technology.

For citation: Córdoba-Tlaxcalteco M.L., Benítez-Guerrero E. A Systematic Literature Review on Vision-Based Human Event Recognition in Smart Classrooms: Identifying Significant Events and Their Applications. *Trudy ISP RAN/Proc. ISP RAS*, vol. 36, issue 1, 2024. pp. 175-198. DOI: 10.15514/ISPRAS-2024-36(1)-11.

Full text: M. L. Córdoba-Tlaxcalteco and E. Benítez-Guerrero. Human Event Recognition in Smart Classrooms Using Computer Vision: A Systematic Literature Review. *Programming and Computer Software*, 2023, Vol. 49, No. 8, pp. 625–642. DOI: 10.1134/S0361768823080066.

Систематический обзор литературы по визуальному распознаванию событий с людьми: выявление значимых событий и их применение

М.Л. Кордова-Тлакскальтеко, ORCID: 0009-0001-8150-8097 <*marcordoba@uv.mx*>
Э. Бенитес-Герреро, ORCID: 0000-0001-5844-4198 <*edbenitez@uv.mx*>

Факультет статистики и информатики университета Веракруса,
Халапа, Веракрус, Мексико.

Аннотация. Область распознавания человеческих событий на основе видения в интеллектуальных средах стала процветающей и успешной дисциплиной, а обширные усилия в области исследований и разработок привели к заметному прогрессу. Этот прогресс не только дал ценную информацию, но также открыл возможность практических применений в различных областях. В этом контексте действия человека, действия, взаимодействия и поведение рассматриваются как события, представляющие интерес в интеллектуальных средах. Однако при сосредоточении внимания на умных классах отсутствие общепризнанного определения «человеческого события» создает серьезную проблему для педагогов, исследователей и разработчиков. Это отсутствие согласия препятствует их способности точно определять и классифицировать конкретные ситуации, имеющие отношение к образовательному контексту. Чтобы решить эту проблему, авторы поставили цель провести систематический обзор литературы о значительных событиях, уделяя особое внимание их применению в вспомогательных технологиях. Обзор включает в себя всесторонний анализ 227 опубликованных документов, охватывающих период с 2012 по 2022 год. Он углубляется в ключевые алгоритмы, методологии и приложения распознавания событий на основе видения в интеллектуальных средах. В качестве основного результата обзор определяет наиболее значимые события, классифицируя их в соответствии с поведением одного человека, взаимодействиями между несколькими людьми или взаимодействиями между объектом и человеком, изучая их практическое применение в образовательном контексте. Документ завершается обсуждением актуальности и практичности распознавания человеческих событий на основе видения в умных классах, особенно в эпоху после COVID.

Ключевые слова: распознавание событий с людьми; умный класс; компьютерное зрение; искусственный интеллект; образовательные технологии.

Для цитирования: Кордова-Тлакскальтеко М. Л., Бенитес-Герреро Э. Систематический обзор литературы по визуальному распознаванию событий с людьми: выявление значимых событий и их применение. Труды ИСП РАН, том 36, вып. 1, 2024 г., стр. 175–198 (на английском языке). DOI: 10.15514/ISPRAS–2024–36(1)–11.

Полный текст: М.Л. Кордова-Тлакскальтеко, Э. Бенитес-Герреро. Распознавание событий с людьми в умных классах на основе машинного зрения: систематический обзор литературы. *Programming and Computer Software*, 2023, т. 49, № 8, стр. 625–642 (на английском языке). DOI: 10.1134/S0361768823080066.

1. Introduction

Human Event Recognition (HER) in Smart Classrooms (SC) involves using computer techniques to identify some human actions, activities, interactions, and behaviors within educational spaces equipped with data acquisition and processing infrastructure. Specifically, video data obtained from cameras in smart classrooms is of particular interest for interpreting educational scenes. This technology enables the detection, learning, recognition, and prediction of learners' and teachers' actions, allowing the system to assess and assist them accordingly [1]. This topic has proven beneficial for classroom management (e.g., automated attendance tracking), learning and teaching support (e.g., detecting social interactions and collaborative learning), and enhancing students' academic performance (e.g., identifying action patterns related to academic achievement) [2][4][5].

Previous reviews have addressed video-based HER (see Table 1). For instance, reference [9] provides an overview of recent vision-based techniques for recognizing human behaviors and surveillance systems. In [10], deep learning methods with automatic feature extraction for vision-based human event recognition are reviewed. The authors of [11] present a comprehensive review of approaches to recognizing and representing human actions through visual data. Reference [12] presents a state-of-the-art review on recognizing suspicious behaviors in surveillance videos, including six different systems. Finally, reference [13] describes major video datasets for human event recognition. It's worth noting that while these works are important as they share common underlying techniques, none of them specifically focuses on SCs. In [243], a conceptual account of SC evolution and its relationship with AI and emerging educational technologies is provided.

This paper aims to analyze the state of the art in vision-based recognition of human events in smart classrooms, with a specific focus on identifying the most significant events. It seeks to provide educators, researchers, and educational technology developers with a comprehensive overview of the topic while also highlighting the lack of consensus on what events are considered the most significant in this context. To achieve this, the paper presents a systematic literature review of published works in the last 10 years. The review covers key concepts and methodologies, drawing from the analysis of 227 documents, and identifies relevant events and their applications in educational settings. By doing so, it aims to address research gaps and identify opportunities for further exploration in this evolving field

The paper is organized as follows. Section 2 provides background information on HER and SCs. Section 3 outlines the systematic review method. Section 4 presents the review's results, including a list of relevant events with references and brief descriptions. Finally, Section 5 concludes the paper.

Table 1. List of similar reviews in the literature

2004	[53]
2005	[54]
2008	[55]
2009	[7]
2010	[56]
2011	[58]
2012	[59], [60], [61]
2013	[62], [63], [64]
2014	[57], [65], [66]
2015	[67], [68], [69]
2016	[70], [71]
2017	[72], [73], [74], [75]
2018	[12], [11]
2019	[10], [77], [78], [79], [80], [81], [82], [83], [84], [85]
2020	[9]

2. Background

This section is organized as follows. First, the discussion focuses on HER and SCs. Next, Computer Vision methods for object extraction are presented. Finally, event understanding from video scenes is examined.

2.1 SC and HER in the Context of Educational Technology

The origins of Smart Classroom (SC) and Human Event Recognition (HER) can be traced back to the late 20th century when computers and the Internet were introduced in educational settings in the

1980s. The 1990s saw the emergence of Computer Supported Collaborative Learning (CSCL), which linked education and computational technology in collaborative settings. The early 2000s witnessed the growth of e-learning, online education, and the Internet of Things (IoT), leading to the establishment of educational spaces with high technological content. Concurrently, advancements in Machine Learning (ML) and Artificial Intelligence (AI) enabled robust and precise object detection and classification based on data, making it applicable in complex situations.

Since 2010, the widespread use of smartphones, mobile devices, and cloud computing has facilitated data collection, storage, processing, and sharing, giving rise to the concept of Smart Environments (SE). The COVID-19 pandemic in 2020 further emphasized the need for technologically assisted educational services in modern society. However, the complexity of the current educational setting remains a significant challenge for HER in SCs [6].

According to the taxonomy in [244], there are four types of SCs: Basic, Interactive, Collaborative, and Immersive. The majority of the reviewed research presented in this article corresponds to Basic SCs, equipped with multimedia equipment and a computer connected to the Internet. In this context, Computer Vision proves advantageous in easily obtaining SC data compared to the use of biometric, ambient, or wearable sensors. Moreover, much of the reviewed research focuses on the psychological, social, or behavioral dimensions of educational experiences. Current trends aim to combine multimodal data acquisition in IoT with data fusion in AI to cater to face-to-face, online, or hybrid educational modalities [85]. References from [231] to [242] review hybrid and sensor-based approaches to HER.

2.2 Computer Vision methods for object extraction

In video data analysis, the first step involves detecting features known as objects, which can be things, people, or combinations of both (see fig. 1). Object Detection (OD), Object Classification (OC), and Object Tracking (OT) are the common processes for extracting features or objects.

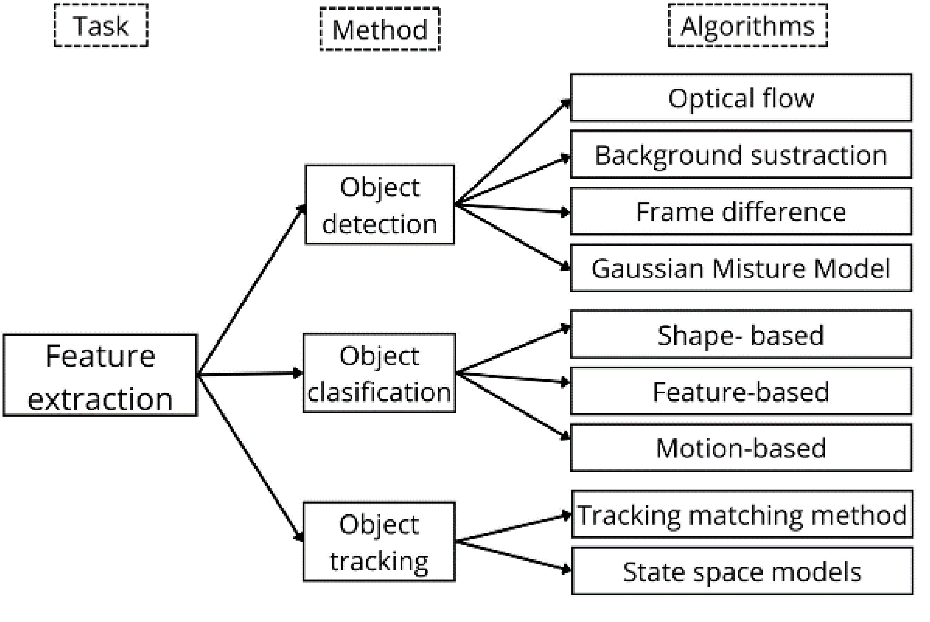

Fig. 1. Schematic representation of methods and algorithms for feature extraction with information from [246], created by the authors

The techniques for feature extraction include Optical Flow, Background Subtraction, Frame Difference, and Gaussian Mixture Model. Optical flow is described in [14], background subtraction in [15], and frame difference in [16]. Gaussian Mixture Model is used to estimate probability distributions, such as Gaussian distribution [17] or a mixture of Gaussian distribution [18], with fast estimation algorithms shown in [19] or [20]. For a more detailed discussion, refer to [246], upon which this section is primarily based.

OC (Object Classification) is the next step, involving shape-based, motion-based, and feature-based methods. Shape-based OC uses geometric properties like height/width ratio, perimeter, and area [21], useful for human figure classification [13, 22-23]. Motion-based classification relies on distinguishing objects based on their motion characteristics, recognizing human movements like walking or running [24-25]. Feature-based classification uses specific frame elements, such as skin color [26], which can also be combined with other descriptors [27].

The final stage is OT (Object Tracking), which creates a track of each object by capturing their locations over time [28]. Tracking Matching Methods find correspondence between object detections in different frames [29-30]. Another category, State Space Models, estimates object state (position, velocity, etc.) using a motion model corrected by incomplete measurements [32], with complete measurements obtained through OD algorithms [32].

2.3 Event understanding from video

Event understanding in video scenes involves interpreting elements based on known context (see fig. 2). It can be data-driven, using supervised and unsupervised machine learning methods like decision trees, KNN, SVMs, HMMs, and Bayesian networks [41]. Unsupervised learning constructs recognition models from unlabeled data using density estimation or clustering methods [35], including graphical models and eigen-decomposition [33-34].

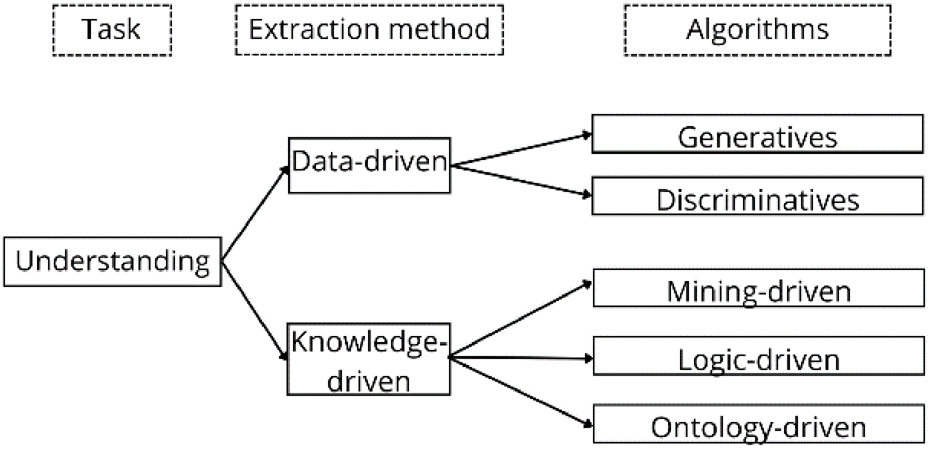

Fig. 2. Schematic representation of methods and algorithms for event understanding with information from [245], created by the authors

Data-driven algorithms in event understanding (fig. 3) can be classified into generative and discriminative methods [35]. Generative methods like Bayes classifiers, Hidden Markov Processes, and Bayesian Networks provide a complete description but require large data volumes for learning parameters. On the other hand, discriminative methods like Deep Learning Neural Networks, SVM, and Nearest Neighbor have lower computational costs but do not fully explain human events. Hybrid methods that combine both approaches have also been proposed [36].

Knowledge-driven understanding in event recognition utilizes formal knowledge [245]. Logical formalisms like Plans Recognition Theory [37-38], and Event Theory [39] are used for HER. Knowledge-driven methods (fig. 4) can be categorized into mining-driven, logic-driven, and ontology-driven approaches [40]. Mining-driven methods learn from pre-defined data to classify behaviors, while logic-driven methods use semantic representations and reasoning mechanisms. Ontology-driven methods, gaining interest in behavior recognition, offer an explicit representation of behavior definitions for broader applicability.

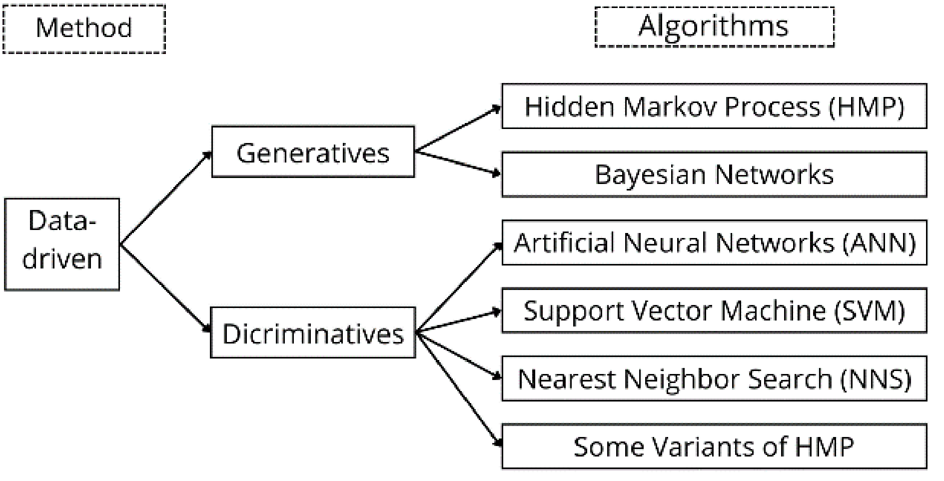

Fig. 3. Schematic representation of algorithms for data driven methods in HER, created by the authors

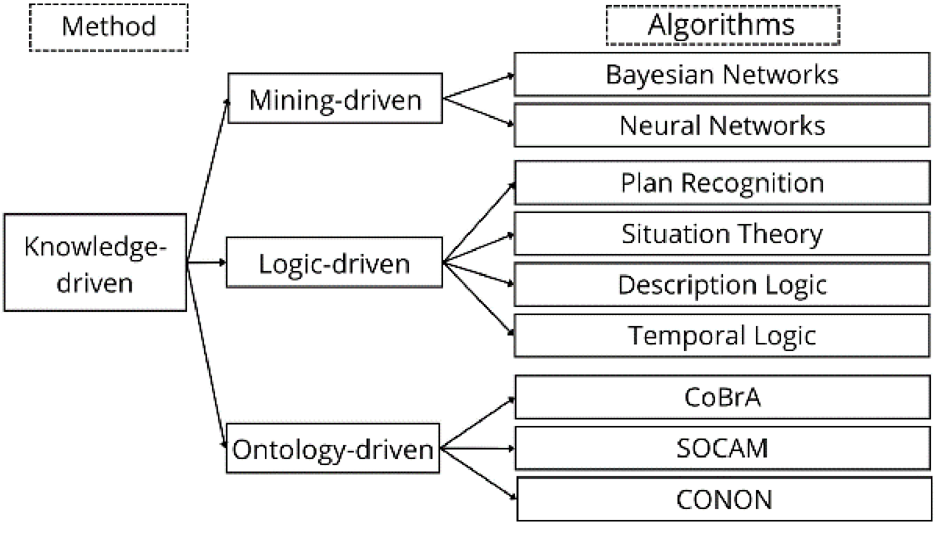

Fig. 4. Schematic representation of algorithms for knowledge-driven methods in HER, created by the authors

Ontologies offer advantages like independence from specific algorithms, promoting portability, interoperability, and reuse of technologies and systems. They have been used to model social interaction in various domains, such as nursing homes, meeting videos, and bank monitoring. Researchers have created a video event ontology for surveillance, leading to its use in scenarios like bank and car park monitoring. While ontologies provide common terms for event definitions, scene interpretation may involve individually preferred algorithms, like rule-based systems and finite-state machines, which may share limitations with logical-based methods.

3. Method of the systematic review

The SLR method used in this study is based on the approach proposed in [52], which involves several stages depicted in Fig. 5.

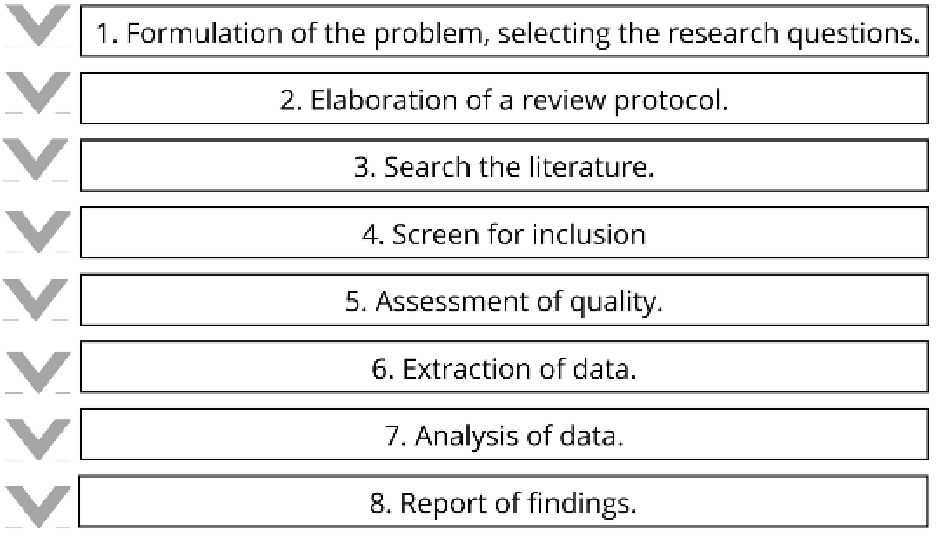

Fig. 5. Schematic representation of the methodology followed on the review, created by the authors

The research process in this study involved several stages:

1. Establishment of Research Questions:
   1.1. RQ1. What research has been conducted on acknowledging events from single-person or non-interacting behavior?
   1.2. RQ2. What research has been conducted on the recognition of events involving multiple-person interactions?
   1.3. RQ3. What research has been conducted on the recognition of events involving people-object interactions?
2. Definition of the Study Plan:

   The plan included determining information sources, inclusion criteria, search strategies, quality assessment criteria, screening procedures, and strategies for data extraction, synthesis, and reporting. The selected digital libraries were ACM, IEEE, Elsevier, and Springer, as they are prominent in the computing field and accessible to the authors. Inclusion criteria covered publications from the last ten years, containing the specified keywords in the title, abstract, or complete document, while reviews or surveys were excluded. The search string used was:

("event" OR "behavior" OR "action" OR "activity" OR "interaction") AND ("recognition" OR "detection" OR "tracking") AND

("smart classroom" OR "classroom") AND ("video" OR "vision").

3. Searching for Relevant Papers:

   The search string was adapted for each source, and relevant papers were sought in the selected digital libraries.

4. Screening and Selection of Papers:

   A two-step process was applied, involving the review of titles and abstracts for inclusion and a full-text review of selected papers.

5. Quality Assessment:

   While quality assessment is important for reviews aiming for generalization, it was not used as a criterion to exclude papers in this study, which sought to discover studies at different quality levels for a more comprehensive overview.

6. Data Extraction:

   Relevant data for answering the research questions was extracted from the selected papers.

7. Analysis

   The gathered data was thoroughly analyzed to draw meaningful conclusions.

8. Reporting:

   The collected data was analyzed, and the resulting findings were comprehensively reported in this paper.

4. *Results of the review*

4.1 Quantitative results

This section presents the results of the SLR conducted in this research, as shown in Fig. 6.

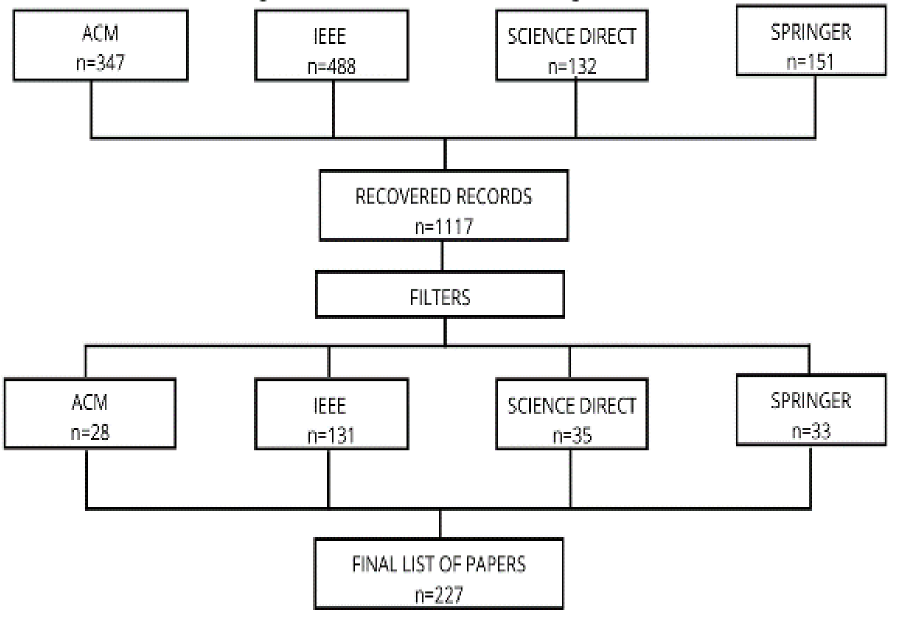

Fig. 6. Schematic representation of the process of selection of documents, created by the authors

Out of 1,117 documents initially selected, 227 papers met the inclusion-exclusion criteria after thorough reviews of titles, abstracts, and content (fig. 7). The list of papers remained unchanged throughout stages 5 and 6.

The papers were grouped by publication year and database. Over the years, the number of papers increased, showing growing interest in the field. Among the 227 documents (fig. 8), IEEE had the most papers (131, 57.7%), followed by Elsevier (35, 15.4%), SPRINGER (33, 14.6%), and ACM (28, 12.3%).

Qualitative analysis results presented by research question.

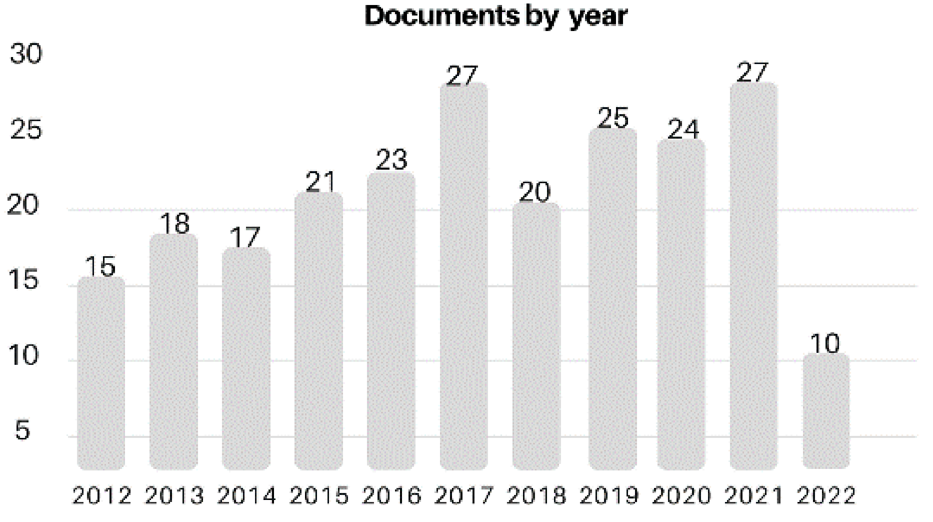

Fig. 7. A total of 227 articles with the search of the keywords, created by the authors

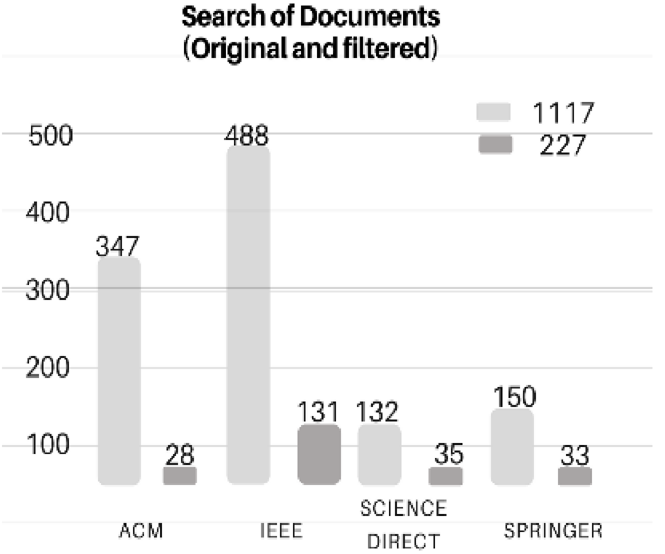

Fig. 8. Total of reviewed articles, by source, created by the authors

4.2 RQ1: Single person or non-interaction events

4.2.1 Event 1. Students being distracted from learning

In the context of educational settings, detecting events related to single person or non-interaction scenarios, such as students being distracted from learning, has been explored using various methods.

Gesture analysis has been utilized to identify boredom and lack of attention in students [92]. Additionally, facial expressions have been studied as indicators of students' feelings, and methods like image recognition and facial muscle tension measurement have been employed to capture facial expressions [93-94, 96].

Eye-gaze and face-gaze analysis have also proven to be important indicators of cognitive engagement among students [97-98]. Researchers have recorded and analyzed human gaze behavior in different scenarios, including conversational gaze and tutoring interactions [98-99, 101-103].

Pose estimation methods have been applied to detect self-absorbed or sleeping students [106-108]. These methods often involve probabilistic and compositional graphical models, but they may encounter challenges in handling errors arising from small body parts in still images [107]. Video pose estimation methods, which incorporate motion information, have been used as well [109]. However, they may have limitations in handling action datasets with larger human motion and appearance variations due to viewpoint changes.

4.2.2 Event 2. Detection of behavior related to developmental disorders

Developmental disorders such as autism and attention disorders, like ADHD, can be detected in the classroom using various computer-based methods.

For autism detection, eye-tracking from computer searching tasks has been employed as an easier, cheaper, and less-obtrusive alternative to fMRI data recording [110-113].

Regarding ADHD diagnosis, facial expression analysis has been a focus of some research works [114-117]. For instance, [114] proposed a methodology using RGBD sensors for diagnostic predictions of ADHD and ASD. Depth capturing cameras, like Microsoft Kinect, have been used to monitor the movement of children in a classroom setting [115]. These cameras allow tracking and analysis of head motion and velocity profiles to measure hyperactivity. Additionally, computerized continuous performance tests are conducted to measure inattention and impulsivity. The test results are then compared to norm data, generating reports for assessment by clinicians.

Overall, these computer-based approaches offer promising avenues for early detection and intervention for developmental disorders in educational environments.

4.2.3 Event 3. Hand-raising gesture detection

Hand-raising is a behavior studied in gaming, Human-Computer Interaction, and classroom settings [119-121]. Detecting hand-raising in a real classroom can be challenging, but vision-based models using video cameras [122-123] and Kinect [124-126] have been developed to address this. Hand gesture recognition involves tracking, representation, and conversion into meaningful commands for human-computer interaction. Techniques include contact-based and vision-based devices [127-129]. Hand gesture recognition relies on detection, tracking, and recognition using visual features like skin color, shape, and motion [131-132]. Model-based methods use tracking to enhance robustness [133-135].

Vision-based hand gesture recognition includes static and dynamic gestures, using classifiers like Hidden Markov Models [137-139]. Learning algorithms vary based on gesture representation, including supervised, unsupervised, and reinforcement learning [122, 140]. For example, static hand gestures are recognized using the Fourier descriptor of a segmentation image [142].

4.3 RQ2. Multiple-person interactions

4.3.1 Event 1. Speaking and talking

Detecting human speaking is important for Human-Computer Interaction and fatigue detection [143]. Lip movement is used to detect speaking, and video-based approaches have been proposed [145]. Methods like lip motion analysis [146-148], Viola-Jones with skin color pixel detection [149], skin-color segmentation with edge projection [150], and fuzzy c-means clustering [151] have been used for lip detection and speech recognition. Feature extraction methods like Log-polar Signature [153] and Haar-like wavelets [154] have been proposed for lip tracking and speech recognition [157].

4.3.2 Event 2. Social interactions

Video-based studies of human sociality focus on workplace settings and classrooms, observing action and sense-making practices in social interactions [158-160]. Social abilities have been linked to academic success, and Proxemics Theory is used to detect human relationships, including non-verbal relations in classrooms [163-165]. Immediacy, which enhances physical and psychological closeness between individuals [168-170], can impact effective communication in educational settings. Teachers' variable physical proximities with students foster effective communication in classrooms [172-173]. Interaction, where learners share perspectives and collaborate, is another important aspect of non-verbal behaviors [174-175]. Learner-centered approaches and collaborative learning are emphasized in education [176], and providing pre-service teachers with video scenes where students interact with each other can support their understanding of these approaches [178]. However, empirical research is needed to validate assumptions regarding video-based cases and student-student interactions in educational settings [177, 180].

4.4 RQ3. People-objects interaction

4.4.1 Unique event. Student engagement detected by interaction with objects

Various works have classified engagement in different ways [181], including student involvement in terms of effort, persistence, and concentration [179], emotional engagement related to feelings of interest or attitude, and cognitive engagement focusing on cognitive effort and strategies [182]. Agentic engagement emphasizes proactive actions taken by students during learning tasks, involving interaction with surroundings or learning objects.

To assess the level of engagement, traditional methods and measures have been introduced [183], such as using student responses as indicators in intelligent tutoring systems [184-185]. Facial movements and features extracted from them have been used [186-187], along with automated measures like response time to problems and quizzes [188-189]. Physiological and neurological measures like electroencephalogram, heart rate, and skin response have also been employed [190-193]. Some studies utilize facial features and SVM classifiers to analyze affective states of students while solving problems [194-195], while others focus on facial expressions and body movements to detect various affective states of engagement [196].

Engagement detection and localization can be performed using face and facial landmark positions in video frames [197], extracting features from small segments of video, and employing regression models or LSTM-based networks for engagement prediction [196]. Open-source utility software like OpenFace has been used to automatically track changes in body posture and facial movements to infer engagement levels through eye gaze and head movement features [198-201].

Several works have classified engagement in different ways [181]. For example, [179] explains student’s involvement in terms of effort, persistence and concentration. Emotional Engagement is related to feelings of interest or attitude towards a particular theme. Cognitive Engagement focuses on allocation of effort, a strategy used, in terms of cognitive effort, for the accomplishment of the task. Other models have introduced another dimension known as Agentic Engagement and emphasize on proactive actions taken by the student for learning a particular task [182]. These tasks sometimes involve interaction of students with surroundings elements or learning objects.

5. Conclusion

The reviewed works show that there are relatively few studies dedicated to Smart Classroom (SC) event recognition [202-203]. While other smart environments like smart homes or smart offices have more extensive research, SC lacks conventions defining relevant events or behaviors [204]. In SC, event recognition is often a step within an application system, where it serves as input for decision-making processes aimed at assisting users [163].

Overall, video-based Human Event Recognition (HER) in SC has shown positive results, but some projects' costs may hinder widespread implementation [163]. Comparatively, other educational developments like E-learning, M-learning, and MOOCs have gained more traction, especially during the COVID-19 pandemic, but they may lack the non-verbal communication found in traditional classrooms [165]. HER has been suggested as a potential solution to address this limitation [6].

References

[1]. K. Ducatel, U. europeenne. Technologies de la societe de l'information, U. europeenne. Institut d'etudes de prospectives technologiques and U. europeenne. Societe de l'information conviviale, Scenarios for ambient intelligence in 2010 (2001).

[2]. D. J. Cook, J. C. Augusto, V. R. Jakkula. Ambient intelligence: Technologies, applications, and opportunities. Pervasive and Mobile Computing 5(4) (2009), 277–298.

[3]. M.K. Saini, N. Goel. How smart are smart classrooms? A review of smart classroom technologies. ACM Computing Surveys (CSUR) 52(6) (2019), 1–28. https://doi.org/10.1145/3365757.

[4]. D. Guinard, M. Fischer, V. Trifa. Sharing using social networks in a composable web of things. In 2010 8th IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops). IEEE, 2010, pp. 702–707. doi:10.1109/percomw.2010.5470524

[5]. V. Radosavljevic, S. Radosavljevic, G. Jelic. Ambient intelligence-based smart classroom model, Interactive Learning Environments (2019). 1–15. https://doi.org/10.1080/10494820.2019.1652836.

[6]. M. Kwet and P. Prinsloo, The 'smart'classroom: a new frontier in the age of the smart university, Teaching in Higher Education 25(4) (2020). 510–526. https://doi.org/10.1080/13562517.2020.1734922.

[7]. J. Candamo, M. Shreve, D.B. Goldgof, D.B. Sapper, R. Kasturi. Understanding transit scenes: A survey on human behavior-recognition algorithms. IEEE transactions on intelligent transportation systems 11(1) (2009), 206–224. 10.1109/TITS.2009.2030963.

[8]. M.A.R. Ahad. Vision and Sensor-Based Human Activity Recognition: Challenges Ahead. In Advancements in Instrumentation and Control in Applied System Applications. IGI Global, 2020, pp. 17–35. DOI:10.4018/978-1-7998-2584-5.

[9]. D.R. Beddiar, B. Nini, M. Sabokrou, A. Hadid. Vision based human activity recognition: a survey. Multimedia Tools and Applications 79(41) (2020), 30509–30555.

[10]. H.-B. Zhang, Y.-X., B. Zhong, Q. Lei, L. Yang, J.-X. Du, D.-S. Chen. A comprehensive survey of vision based human action recognition methods. Sensors 19(5) (2019), 1005.

[11]. Y. Kong, Y. Fu. Human action recognition and prediction: A survey. arXiv preprint arXiv:1806.11230 (2018).

[12]. R.K. Tripathi, A.S. Jalal, S.C. Agrawal. Suspicious human activity recognition: a review. Artificial Intelligence Review 50(2) (2018), 283–339. https://doi.org/10.1007/s10462-017-9545-7.

[13]. I. Jegham, A.B. Khalifa, I. Alouani, M.A. Mahjoub. Vision-based human action recognition: An overview and real world challenges. Forensic Science International: Digital Investigation 32 (2020), 200901. https://doi.org/10.1016/j.fsidi.2019.200901.

[14]. F. Paredes-Valles, K.Y. Scheper, G.C. De Croon. Unsupervised learning of a hierarchical spiking neural network for optical flow estimation: From events to global motion perception. IEEE transactions on pattern analysis and machine intelligence 42(8) (2019), 2051–2064.

[15]. B. Garcia-Garcia, T. Bouwmans, A.J.R. Silva. Background subtraction in real applications: Challenges, current models and future directions. Computer Science Review 35 (2020), 100204.

[16]. M. Ahmad, I. Ahmed, K. Ullah, I. Khan, A. Adnan. Robust background subtraction based person's counting from overhead view. In 2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), IEEE, 2018, pp. 746–752.
[17]. Y. Fan, G. Wen, D. Li, S. Qiu, M.D. Levine, F. Xiao. Video anomaly detection and localization via Gaussian mixture fully convolutional variational autoencoder. Computer Vision and Image Understanding (2020), 102920. https://doi.org/10.1016/j.cviu.2020.102920.
[18]. J. Cho, Y. Jung, D.-S. Kim, S. Lee, Y. Jung. Moving object detection based on optical flow estimation and a Gaussian mixture model for advanced driver assistance systems. Sensors 19(14) (2019), 3217.
[19]. H. Yang, S. Qu. Real-time vehicle detection and counting in complex traffic scenes using background subtraction model with low-rank decomposition. IET Intelligent Transport Systems 12(1) (2018), 75-85.
[20]. I. Martins, P. Carvalho, L. Corte-Real, J.L. Alba-Castro. Bmog: boosted gaussian mixture model with controlled complexity for background subtraction. Pattern Analysis and Applications 21(3) (2018), 641–654.
[21]. N. Zerrouki, F. Harrou, Y. Sun, A. Houacine. Visionbased human action classification using adaptive boosting algorithm. IEEE Sensors Journal 18(12) (2018), 5115–5121.
[22]. L. Boregowda, A. Rajagopal. Object classification in video data. Google Patents, 2007, US Patent App.11/227,505.
[23]. N.D. Bird, O. Masoud, N.P. Papanikolopoulos, A. Isaacs. Detection of loitering individuals in public transportation areas. IEEE Transactions on intelligent transportation systems 6(2) (2005), 167–177. 10.1109/TITS.2005.848370.
[24]. R. Zhang, C. Vogler, D. Metaxas. Human gait recognition at sagittal plane. Image and vision computing 25(3) (2007), 321–330. https://doi.org/10.1016/j.imavis.2005.10.007.
[25]. V. Kukreja, D. Kumar, A. Kaur. Deep learning in Human Gait Recognition: An Overview. In 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), IEEE, 2021, pp. 9–13.
[26]. U.Sevik, E. Karakullukcu, T. Berber, Y. Akbas, S. Turkyılmaz. Automatic classification of skin burn colour images using texture-based feature extraction. IET Image Processing 13(11) (2019), 2018–2028.
[27]. M.J. Taylor and T. Morris, Adaptive skin segmentation via feature-based face detection. In Real-Time Image and Video Processing 2014, Vol. 9139, International Society for Optics and Photonics, 2014, p. 91390P. https://doi.org/10.1117/12.2052003.
[28]. S.A. Rodriguez, V. Fremont, P. Bonnifait, V. Cherfaoui. An embedded multi-modal system for object localization and tracking. IEEE Intelligent Transportation Systems Magazine 4(4) (2012), 42–53. 10.1109/MITS.2012.2217855.
[29]. Y. Lu, C. Lu, C.-K. Tang. Online video object detection using association LSTM. In Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2344–2352. 10.1109/ICCV.2017.257.
[30]. R. Messing, C. Pal, H. Kautz. Activity recognition using the velocity histories of tracked keypoints. In 2009 IEEE 12th international conference on computer vision, IEEE, 2009, pp. 104–111. 10.1109/ICCV.2009.5459154.
[31]. I. Leichter, M. Lindenbaum, E. Rivlin. Mean shift tracking with multiple reference color histograms. Computer Vision and Image Understanding 114(3) (2010), 400–408. https://doi.org/10.1016/j.cviu.2009.12.006.
[32]. Y. Wu, J. Lim, M.-H. Yang. Online object tracking: A benchmark. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 2411–2418. 10.1109/TPAMI.2014.2388226.
[33]. E. Alpaydin. Introduction to machine learning. MIT press, 2020. https://doi.org/10.1016/j.neuroimage.2010.11.004.
[34]. Y. Li. On incremental and robust subspace learning. Pattern recognition 37(7) (2004), 1509–1518. https://doi.org/10.1016/j.patcog.2003.11.010.
[35]. F. Mairesse, M. Gasic, F. Jurcicek, S. Keizer, B. Thomson, K. Yu, S. Young. Spoken language understanding from unaligned data using discriminative classification models. In 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2009, pp. 4749–4752. 10.1109/ICASSP.2009.4960692.
[36]. K.Z. Haigh, L.M. Kiff, G. Ho. The independent lifestyle assistant: Lessons learned. Assistive Technology 18(1) (2006), 87–106. https://doi.org/10.1080/10400435.2006.10131909.

[37]. L. Chen, I. Khalil. Activity recognition: Approaches, practices and trends. In Activity Recognition in Pervasive Intelligent Environments, Springer, 2011, pp. 1–31. https://doi.org/10.2991/978-94-91216-05-3_1.
[38]. L. Chen, C.D. Nugent. Human Activity Recognition and Behaviour Analysis. Springer, 2019.
[39]. B. Bouchard, S. Giroux, A. Bouzouane. A smart home agent for plan recognition of cognitively-impaired patients. JCP 1(5) (2006), 53–62.
[40]. L. Chen, C.D. Nugent, H. Wang. A knowledge-driven approach to activity recognition in smart homes. IEEE Transactions on Knowledge and Data Engineering 24(6) (2011), 961–974. 10.1109/TKDE.2011.51.
[41]. R.L. Granada, R.F. Pereira, J. Monteiro, R.C. Barros, D. Ruiz, F. Meneguzzi. Hybrid activity and plan recognition for video streams. In Workshops at the Thirty-First AAAI Conference on Artificial Intelligence, 2017.
[42]. H. Kautz, O. Etzioni, D. Fox, D. Weld, L. Shastri. Foundations of assisted cognition systems. University of Washington, Computer Science Department, Technical Report, Tech. Rep (2003).
[43]. E.S. Chen, G.B. Melton, M.E. Engelstad, I.N. Sarkar. Standardizing clinical document names using the HL7/LOINC document ontology and LOINC codes. In AMIA Annual Symposium Proceedings, Vol. 2010, American Medical Informatics Association, 2010, p. 101.
[44]. A. Hakeem, M. Shah. Multiple agent event detection and representation in videos. In AAAI, 2005, pp. 89–94.
[45]. B. Georis, M. Maziere, F. Bremond, M. Thonnat. A video interpretation platform applied to bank agency monitoring (2004). 10.1049/ic:20040097.
[46]. J.C. SanMiguel, J.M. Martinez, A. Garcia. An ontology for event detection and its application in surveillance video. In 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, IEEE, 2009, pp. 220–225. 10.1109/AVSS.2009.28.
[47]. A.R. Francois, R. Nevatia, J. Hobbs, R.C. Bolles, J.R. Smith. VERL: an ontology framework for representing and annotating video events. IEEE multimedia 12(4) (2005), 76–86. 10.1109/AVSS.2009.28.
[48]. U. Akdemir, P. Turaga, R. Chellappa. An ontology based approach for activity recognition from video. In Proceedings of the 16th ACM international conference on Multimedia, 2008, pp. 709–712. https://doi.org/10.1145/1459359.1459466.
[49]. B. Yao, H. Hagras, M.J. Alhaddad, D. Alghazzawi. A fuzzy logic-based system for the automation of human behavior recognition using machine vision in intelligent environments. Soft Computing 19(2) (2015), 499–506. DOI10.1007/s00500-014-1270-4.
[50]. N. Ikizler, D. Forsyth. Searching video for complex activities with finite state models. In 2007 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2007, pp. 1–8.
[51]. L. Ballan, M. Bertini, A. Del Bimbo, L. Seidenari, G. Serra. Event detection and recognition for semantic annotation of video. Multimedia tools and applications 51(1) (2011), 279–302. http://hdl.handle.net/11380/979935.
[52]. Y. Xiao, M. Watson. Guidance on conducting a systematic literature review. Journal of Planning Education and Research 39(1) (2019), 93–112.
[53]. W. Hu, T. Tan, L. Wang, S. Maybank. A survey on visual surveillance of object motion and behaviors. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 34(3) (2004), 334–352. 10.1109/TSMCC.2004.829274.
[54]. M. Valera, S.A. Velastin. Intelligent distributed surveillance systems: a review. IEE Proceedings-Vision, Image and Signal Processing 152(2) (2005), 192–204. 10.1049/ip-vis:20041147.
[55]. P. Turaga, R. Chellappa, V.S. Subrahmanian, O. Udrea. Machine recognition of human activities: A survey. IEEE Transactions on Circuits and Systems for Video technology 18(11) (2008), 1473–1488. 10.1109/TCSVT.2008.2005594.
[56]. R. Poppe. A survey on vision-based human action recognition. Image and vision computing 28(6) (2010), 976–990. https://doi.org/10.1016/j.imavis.2009.11.014.
[57]. J.K. Aggarwal, L. Xia. Human activity recognition from 3d data: A review. Pattern Recognition Letters 48 (2014), 70– 80. https://doi.org/10.1016/j.patrec.2014.04.011.
[58]. J.K. Aggarwal, M.S. Ryoo. Human activity analysis: A review. ACM Computing Surveys (CSUR) 43(3) (2011), 1–43. https://doi.org/10.1145/1922649.1922653.
[59]. A.A. Chaaraoui, P. Climent-Perez, F. Florez-Revuelta. A review on vision techniques applied to human behaviour analysis for ambient-assisted living. Expert Systems with Applications 39(12) (2012), 10873–10888. https://doi.org/10.1016/j.eswa.2012.03.005.

[60]. O.P. Popoola, K. Wang. Video-based abnormal human behavior recognition—A review. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 42(6) (2012), 865–878. 10.1109/TSMCC.2011.2178594.
[61]. A.A. Sodemann, M.P. Ross, B.J. Borghetti. A review of anomaly detection in automated surveillance. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 42(6) (2012), 1257–1272. 10.1109/TSMCC.2012.2215319.
[62]. P.V.K. Borges, N. Conci, A. Cavallaro. Video-based human behavior understanding: A survey. IEEE transactions on circuits and systems for video technology 23(11) (2013),1993–2008. 10.1109/TCSVT.2013.2270402.
[63]. J.M. Chaquet, E.J. Carmona, A. Fernandez-Caballero. A survey of video datasets for human action and activity recognition. Computer Vision and Image Understanding 117(6) (2013), 633–659. https://doi.org/10.1016/j.cviu.2013.01.013.
[64]. S. Vishwakarma, A. Agrawal. A survey on activity recognition and behavior understanding in video surveillance. The Visual Computer 29(10) (2013), 983–1009. https://doi.org/10.1007/s00371-012-0752-6.
[65]. G. Guo, A. Lai. A survey on still image based human action recognition. Pattern Recognition 47(10) (2014), 3343–3361. https://doi.org/10.1016/j.patcog.2014.04.018.
[66]. S.A. Lowe, G. OLaighin. Monitoring human health behaviour in one’s living environment: a technological review. Medical engineering & physics 36(2) (2014), 147–168. https://doi.org/10.1016/j.medengphy.2013.11.010.
[67]. M. Amiribesheli, A. Benmansour, A. Bouchachia. A review of smart homes in healthcare. Journal of Ambient Intelligence and Humanized Computing 6(4) (2015), 495–517. https://doi.org/10.1007/s12652-015-0270-2.
[68]. P. Hurney, P. Waldron, F. Morgan, E. Jones, M. Glavin. Review of pedestrian detection techniques in automotive farinfrared video. IET intelligent transport systems 9(8) (2015), 824–832. DOI:10.1049/iet-its.2014.0236.
[69]. M. Ziaeefard, R. Bergevin. Semantic human activity recognition: A literature review. Pattern Recognition 48(8) (2015), 2329–2345. https://doi.org/10.1016/j.patcog.2015.03.006.
[70]. M. Ramezani, F. Yaghmaee. A review on human action analysis in videos for retrieval applications. Artificial Intelligence Review 46(4) (2016), 485–514. DOI10.1007/s10462-016-9473-y.
[71]. T. Subetha, S. Chitrakala. A survey on human activity recognition from videos. In 2016 International Conference on Information Communication and Embedded Systems (ICICES), IEEE, 2016, pp. 1–7. 10.1109/ICICES.2016.7518920.
[72]. J. Mahata, A. Phadikar. Recent advances in human behaviour understanding: A survey. In 2017 Devices for Integrated Circuit (DevIC), IEEE, 2017, pp. 751–755. 10.1109/DEVIC.2017.8074052.
[73]. S. Rashmi, S. Bhat, V. Sushmitha, Evaluation of human action recognition techniques intended for video analytics. In 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon), IEEE, 2017, pp. 357–362.0.1109/SmartTechCon.2017.8358396.
[74]. K. Rohit, K. Mistree, J. Lavji. A review on abnormal crowd behavior detection. In 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), IEEE, 2017, pp. 1–3. 10.1109/ICIIECS.2017.8275999.
[75]. A.B. Sargano, P. Angelov, Z. Habib. A comprehensive review on handcrafted and learning-based action representation approaches for human activity recognition, applied sciences 7(1) (2017), 110. https://doi.org/10.3390/app7010110.
[76]. M. Lussier, M. Lavoie, S. Giroux, C. Consel, M. Guay, J. Macoir, C. Hudon, D. Lorrain, L. Talbot, F. Langlois et al. Early detection of mild cognitive impairment with in-home monitoring sensor technologies using functional measures: A systematic review. IEEE journal of biomedical and health informatics 23(2) (2018), 838–847. 10.1109/JBHI.2018.2834317.
[77]. C. Dhiman, D.K. Vishwakarma. A review of state-of-theart techniques for abnormal human activity recognition. Engineering Applications of Artificial Intelligence 77 (2019), 21–45. https://doi.org/10.1016/j.engappai.2018.08.014.
[78]. M. Fahim, A. Sillitti, Anomaly Detection. Analysis and Prediction Techniques in IoT Environment: A Systematic Literature Review. IEEE Access 7 (2019), 81664–81681. 10.1109/ACCESS.2019.2921912.
[79]. R. Iguernaissi, D. Merad, K. Aziz, P. Drap. People tracking in multi-camera systems: a review. Multimedia Tools and Applications 78(8) (2019), 10773–10793. https://doi.org/10.1007/s11042-018-6638-5.

[80]. A. Lentzas, D. Vrakas. Non-intrusive human activity recognition and abnormal behavior detection on elderly people: a review. Artificial Intelligence Review (2019), 1–47. https://doi.org/10.1007/s10462-019-09724-5.
[81]. S. Nigam, R. Singh, A. Misra, A Review of Computational Approaches for Human Behavior Detection. Archives of Computational Methods in Engineering 26(4) (2019), 831–863. https://doi.org/10.1007/s11831-018-9270-7.
[82]. T. Sikandar, K.H. Ghazali, M.F. Rabbi, ATM crime detection using image processing integrated video surveillance: a systematic review. Multimedia Systems 25(3) (2019), 229–251. https://doi.org/10.1007/s00530-018-0599-4.
[83]. T. Singh, D.K. Vishwakarma. Video benchmarks of human action datasets: a review. Artificial Intelligence Review 52(2) (2019), 1107–1154. https://doi.org/10.1007/s10462-018-9651-1.
[84]. R.K. Tripathi, A.S. Jalal, S.C. Agrawal. Abandoned or removed object detection from visual surveillance: a review. Multimedia Tools and Applications 78(6) (2019), 7585–7620. https://doi.org/10.1007/s11042-018-6472-9.
[85]. J. Wang, Y. Chen, S. Hao, X. Peng, L. Hu. Deep learning for sensor-based activity recognition: A survey. Pattern Recognition Letters 119 (2019), 3–11. https://doi.org/10.1016/j.patrec.2018.02.010.
[86]. U. Bakar, H. Ghayvat, S. Hasanm, S. Mukhopadhyay. Activity and anomaly detection in smart home: A survey. In Next Generation Sensors and Systems, Springer, 2016, pp. 191–220. https://doi.org/10.1007/978-3-319-21671-3_9.
[87]. C. Zhang, Q.-S. Jia. A review of occupant behavior models in residential building: Sensing, modeling, and prediction. In 2016 Chinese Control and Decision Conference (CCDC), IEEE, 2016, pp. 2032–2037. 10.1109/CCDC.2016.7531318.
[88]. A.S. Al-Shamayleh, R. Ahmad, M.A. Abushariah, K.A. Alam, N. Jomhari, A systematic literature review on vision based gesture recognition techniques. Multimedia Tools and Applications 77(21) (2018), 28121–28184. https://doi.org/10.1007/s11042-018-5971-z.
[89]. A.B. Mabrouk, E. Zagrouba. Abnormal behavior recognition for intelligent video surveillance systems: A review. Expert Systems with Applications 91 (2018), 480–491. https://doi.org/10.1016/j.eswa.2017.09.029.
[90]. G. Sreenu, M.S. Durai. Intelligent video surveillance: a review through deep learning techniques for crowd analysis. Journal of Big Data 6(1) (2019), 48. https://doi.org/10.1186/s40537-019-0212-5.
[91]. S.M. Al-Zoubi, M.A.B. Younes. Low academic achievement: causes and results, Theory and Practice. In Language Studies 5(11) (2015), 2262–2268. http://dx.doi.org/10.17507/tpls.0511.09.
[92]. H. Kim, S. Lee, Y. Kim, S. Lee, D. Lee, J. Ju, H. Myung. Weighted joint-based human behavior recognition algorithm using only depth information for low-cost intelligent video-surveillance system. Expert systems with Applications 45 (2016), 131–141. https://www.sciencedirect.com/science/article/abs/pii/S0957417415006648.
[93]. R.L. Mandryk, M.S. Atkins, A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. International journal of humancomputer studies 65(4) (2007), 329–347. https://doi.org/10.1016/j.ijhcs.2006.11.011.
[94]. K.D. Sidney, S.D. Craig, B. Gholson, S. Franklin, R. Picard, A.C. Graesser. Integrating affect sensors in an intelligent tutoring system. In Affective Interactions: The Computer in the Affective Loop Workshop at, 2005, pp. 7–13.
[95]. M. C. Su, C. T. Cheng, M. C. Chang, Y. Z. Hsieh (2021). A video analytic in-class student concentration monitoring system. IEEE Transactions on Consumer Electronics, 67(4), 294-304.
[96]. H. Admoni, B. Scassellati. Social eye gaze in humanrobot interaction: a review. Journal of Human-Robot Interaction 6(1) (2017), 25–63. https://doi.org/10.5898/JHRI.6.1.Admoni.
[97]. K. Jokinen, H. Furukawa, M. Nishida, S. Yamamoto. Gaze and turn-taking behavior in casual conversational interactions. ACM Transactions on Interactive Intelligent Systems (TiiS) 3(2) (2013), 1–30. https://doi.org/10.1145/2499474.2499481.
[98]. S. Andrist, B. Mutlu, A. Tapus. Look like me: matching robot personality via gaze to increase motivation. In Proceedings of the 33rd annual ACM conference on human factors in computing systems, 2015, pp. 3603–3612. https://doi.org/10.1145/2702123.2702592.
[99]. C.T. Ishi, C. Liu, H. Ishiguro, N. Hagita. Head motion during dialogue speech and nod timing control in humanoid robots. In 2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE, 2010, pp. 293–300. 10.1109/HRI.2010.5453183.
[100]. K. Otsuka, Y. Takemae, J. Yamato. A probabilistic inference of multiparty-conversation structure based on Markovswitching models of gaze patterns, head directions, and utterances. In Proceedings of

the 7th international conference on Multimodal interfaces, 2005, pp. 191–198. https://doi.org/10.1145/1088463.1088497.
[101]. C.-M. Huang, B. Mutlu. Learning-based modeling of multimodal behaviors for humanlike robots. In 2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE, 2014, pp. 57–64. https://doi.org/10.1145/2559636.2559668.
[102]. H. Admoni, A. Dragan, S.S. Srinivasa, B. Scassellati. Deliberate delays during robot-to-human handovers improve compliance with gaze communication. In Proceedings of the 2014 ACM/IEEE international conference on Humanrobot interaction, 2014, pp. 49–56. https://doi.org/10.1145/2559636.2559682
[103]. C. Rich, B. Ponsler, A. Holroyd, C.L. Sidner. Recognizing engagement in human-robot interaction. In 2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE, 2010, pp. 375–382. 10.1109/HRI.2010.5453163.
[104]. K. Sakita, K. Ogawara, S. Murakami, K. Kawamura, K. Ikeuchi. Flexible cooperation between human and robot by interpreting human intention from gaze information. In 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566), Vol. 1, IEEE, 2004, pp. 846–851. 10.1109/IROS.2004.1389458.
[105]. S. Andrist, B. Mutlu, M. Gleicher. Conversational gaze aversion for virtual agents. In International Workshop on Intelligent Virtual Agents, Springer, 2013, pp. 249–262. https://doi.org/10.1007/978-3-642-40415-3_22.
[106]. B. Xiaohan Nie, C. Xiong, S.-C. Zhu. Joint action recognition and pose estimation from video. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2015, pp. 1293–1301.
[107]. D.C. Luvizon, D. Picard, H. Tabia. 2d/3d pose estimation and action recognition using multitask deep learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5137–5146. arXiv:1802.09232v2.
[108]. C. Gao, S. Ye, H. Tian, Y. Yan (2021). Multi-scale single-stage pose detection with adaptive sample training in the classroom scene. Knowledge-Based Systems, 222, 107008.
[109]. Y. Yang, D. Ramanan. Articulated human detection with flexible mixtures of parts. IEEE transactions on pattern analysis and machine intelligence 35(12) (2012), 2878–2890.10.1109/TPAMI.2012.261.
[110]. A. Cherian, J. Mairal, K. Alahari, C. Schmid. Mixing body-part sequences for human pose estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2353–2360. 10.1109/CVPR.2014.302.
[111]. J. Kang, X. Han, J. Song, Z. Niu, X. Li (2020). The identification of children with autism spectrum disorder by SVM approach on EEG and eye-tracking data. Computers in biology and medicine, 120, 103722.
[112]. N. V. Valtakari, I. T. Hooge, C. Viktorsson, P. Nyström, T. Falck-Ytter, R. S. Hessels (2021). Eye tracking in human interaction: Possibilities and limitations. Behavior Research Methods, 1-17.
[113]. V. Yaneva, S. Eraslan, Y. Yesilada, R. Mitkov (2020). Detecting high-functioning autism in adults using eye tracking and machine learning. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28(6), 1254-1261.
[114]. V. Yaneva, L.A. Ha, S. Eraslan, Y. Yesilada, R. Mitkov. Detecting autism based on eye-tracking data from web searching tasks. In Proceedings of the Internet of Accessible Things, 2018, pp. 1–10. https://doi.org/10.1145/3192714.3192819.
[115]. S. Jaiswal, M. F. Valstar, A. Gillott, & D. Daley (2017, May). Automatic detection of ADHD and ASD from expressive behaviour in RGBD data. In 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017) (pp. 762-769). IEEE.
[116]. R. Gao, K. Deng, & M. Xie (2022, October). Deep learning-assisted ADHD diagnosis. In Proceedings of the 3rd International Symposium on Artificial Intelligence for Medicine Sciences (pp. 142-147).
[117]. J. Singh, G. Goyal (2021). Decoding depressive disorder using computer vision. Multimedia Tools and Applications, 80, 8189-8212.
[118]. A. Hernandez-Vela, M. Reyes, L. Igual, J. Moya, V. Violant, S. Escalera. ADHD indicators modelling based on Dynamic Time Warping from RGBD data: A feasibility study. In VI CVC Workshop on the progress of Research & Development, Barcelona, Computer Vision Center, Citeseer, 2011, pp. 59–62.
[119]. S.S. Rautaray, A. Agrawal. Vision based hand gesture recognition for human computer interaction: a survey. Artificial intelligence review 43(1) (2015), 1–54. 10.1007/s10462-012-9356-9.

[120]. R. C. Hsu, P. C. Su, J. L. Hsu, C. Y. Wang. (2020, October). Real-time interaction system of human-robot with hand gestures. In 2020 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE) (pp. 396-398). IEEE.

[121]. H.T. Binh, N.Q. Trung, H.-A.T. Nguyen, B.T. Duy. Detecting Student Engagement in Classrooms for Intelligent Tutoring Systems. In 2019 23rd International Computer Science and Engineering Conference (ICSEC), IEEE, pp. 145–149. 10.1109/ICSEC47112.2019.8974739.

[122]. C.-Y. Fang, M.-H. Kuo, G.-C. Lee, S.-W. Chen. Student gesture recognition system in classroom 2.0. In 14th IASTED International Conference on Computers and Advanced Technology in Education, CATE 2011, 2011, pp. 290–297. 10.2316/P.2011.734-010.

[123]. T.S. Nazare, M. Ponti. Hand-raising gesture detection with Lienhart-Maydt method in videoconference and distance learning. In Iberoamerican Congress on Pattern Recognition, Springer, 2013, pp. 512–519. https://doi.org/10.1007/978-3-642-41827-3_64.

[124]. B. Hariharan, S. Padmini, U. Gopalakrishnan. Gesture recognition using Kinect in a virtual classroom environment. In 2014 Fourth International Conference on Digital Information and Communication Technology and its Applications (DICTAP), IEEE, 2014, pp. 118–124. 10.1109/DICTAP.2014.6821668.

[125]. S. Salous, J. Newton, L. Leroy, S. Chendeb. Gestural Recognition by a Four Kinect Module in a CAVE "Le SAS". In RoCHI, 2015, pp. 111–114.

[126]. S. Kapgate, P. Sahu, M. Das, D. Gupta. (2022, May). Human Following Robot using Kinect in Embedded Platform. In 2022 1st International Conference on the Paradigm Shifts in Communication, Embedded Systems, Machine Learning and Signal Processing (PCEMS) (pp. 119-123). IEEE.

[127]. S. Corera, N. Krishnarajah. Capturing hand gesture movement: a survey on tools, techniques and logical considerations. Proceedings of chi sparks (2011).

[128]. V. A. Shanthakumar, C. Peng, J. Hansberger, L. Cao, S. Meacham, V. Blakely (2020). Design and evaluation of a hand gesture recognition approach for real-time interactions. Multimedia Tools and Applications, 79, 17707-17730.

[129]. T. H. Tsai, C. C. Huang, K. L. Zhang (2020). Design of hand gesture recognition system for human-computer interaction. Multimedia Tools and Applications, 79, 5989-6007.

[130]. K.H. Nguyen. Method and apparatus for real-time gesture recognition. Google Patents, 2000, US Patent 6,072,494.

[131]. M. Cote, P. Payeur, G. Comeau. Comparative study of adaptive segmentation techniques for gesture analysis in unconstrained environments. In Proceedings of the 2006 IEEE International Workshop on Imaging Systems and Techniques (IST 2006), IEEE, 2006, pp. 28–33. 10.1109/IST.2006.1650770.

[132]. O. Köpüklü, A. Gunduz, N. Kose, G. Rigoll (2019, May). Real-time hand gesture detection and classification using convolutional neural networks. In 2019 14th IEEE international conference on automatic face & gesture recognition (FG 2019) (pp. 1-8). IEEE.

[133]. D. Liu, L. Zhang, T. Luo, L. Tao, Y. Wu (2020). Towards interpretable and robust hand detection via pixel-wise prediction. Pattern Recognition, 105, 107202.

[134]. Z. Sun, J. Chen, M. Mukherjee, C. Liang, W. Ruan, Z. Pan (2022). Online multiple object tracking based on fusing global and partial features. Neurocomputing, 470, 190-203.

[135]. L. Huang, B. Zhang, Z. Guo, Y. Xiao, Z. Cao, J. Yuan (2021). Survey on depth and RGB image-based 3D hand shape and pose estimation. Virtual Reality & Intelligent Hardware, 3(3), 207-234.

[136]. H. Tang, H. Liu, W. Xiao, N. Sebe (2019). Fast and robust dynamic hand gesture recognition via key frames extraction and feature fusion. Neurocomputing, 331, 424-433.

[137]. T. Song, H. Zhao, Z. Liu, H. Liu, Y. Hu, D. Sun (2021). Intelligent human hand gesture recognition by local-global fusing quality-aware features. Future Generation Computer Systems, 115, 298-303.

[138]. T. L. Dang, S. D. Tran, T. H. Nguyen, S. Kim, N. Monet (2022). An improved hand gesture recognition system using keypoints and hand bounding boxes. Array, 16, 100251.

[139]. N. Aloysius, M. Geetha (2020). Understanding vision-based continuous sign language recognition. Multimedia Tools and Applications, 79(31-32), 22177-22209.

[140]. S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Mekhtiche, M. S. Hossain (2019). Deep Learning for EEG motor imagery classification based on multi-layer CNNs feature fusion. Future Generation computer systems, 101, 542-554.

[141]. Z. Li, D. Lin, X. Tang. Nonparametric discriminant analysis for face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(4) (2009), 755–761. 10.1109/TPAMI.2008.174.

[142]. Y. Ren, F. Zhang. Hand gesture recognition based on MEB-SVM. In 2009 International Conference on Embedded Software and Systems, IEEE, 2009, pp. 344–349. 10.1109/ICESS.2009.21.

[143]. S. Afroze, M. M. Hoque (2020, December). Towards lip motion based speaking mode detection using residual neural networks. In International Conference on Soft Computing and Pattern Recognition (pp. 166-175). Cham: Springer International Publishing.
[144]. S. Afroze, M. M. Hoque. Talking vs Non-Talking: A Vision Based Approach to Detect Human Speaking Mode. In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), IEEE, 2019, pp. 1–6.
[145]. A. Wuerkaixi, Y. Zhang, Z. Duan, C. Zhang (2022, August). Rethinking audio-visual synchronization for active speaker detection. In 2022 IEEE 32nd International Workshop on Machine Learning for Signal Processing (MLSP) (pp. 01-06). IEEE.
[146]. F. Haider, S. Al Moubayed. Towards speaker detection using lips movements for human machine multiparty dialogue. FONETIK 2012 (2012), 117.
[147]. W-N. Lie, H.-C. Hsieh. Lips detection by morphological image processing. In ICSP’98. 1998 Fourth International Conference on Signal Processing (Cat. No. 98TH8344), Vol. 2, IEEE, 1998, pp. 1084–1087.
[148]. M. Bendris, D. Charlet, G. Chollet, Lip activity detection for talking faces classification in TV-Content. In International Conference on Machine Vision, 2010, pp. 187–190.
[149]. I. Khan, H. Abdullah, M.S.B. Zainal. Efficient eyes and mouth detection algorithm using combination of Viola Jones and skin color pixel detection. International Journal of Engineering 3(4) (2013), 8269.
[150]. H.-Y. Huang, Y.-C. Lin. An efficient mouth detection based on face localization and edge projection. International Journal of Computer Theory and Engineering 5(3) (2013),514.
[151]. T. Azim, M.A. Jaffar, A.M. Mirza, Fully automated real time fatigue detection of drivers through fuzzy expert systems. Applied Soft Computing 18 (2014), 25–38.
[152]. R. Navarathna, P. Lucey, D. Dean, C. Fookes, S. Sridharan. Lip detection for audio-visual speech recognition in-car environment. In 10th International Conference on Information Science, Signal Processing and their Applications (ISSPA 2010), IEEE, 2010, pp. 598–601.
[153]. N. Eveno, A. Caplier, P.-Y. Coulon. Accurate and quasiautomatic lip tracking, IEEE Transactions on circuits and systems for video technology 14(5) (2004), 706–715.
[154]. C. Bouvier, A. Benoit, A. Caplier, P.-Y. Coulon. Open or closed mouth state detection: static supervised classification based on log-polar signature. In International Conference on Advanced Concepts for Intelligent Vision Systems, Springer, 2008, pp. 1093–1102.
[155]. K. Saenko, K. Livescu, M. Siracusa, K. Wilson, J. Glass, T. Darrell. Visual speech recognition with loosely synchronized feature streams. In Tenth IEEE International Conference on Computer Vision (ICCV’05) Volume 1, Vol. 2, IEEE, 2005, pp. 1424–1431.
[156]. M.I. Faraj, J. Bigun, Person verification by lip-motion. In 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW’06), IEEE, 2006, pp. 37–37.
[157]. P.D. Polur, G.E. Miller. Experiments with fast Fourier transform, linear predictive and cepstral coefficients in dysarthric speech recognition algorithms using hidden Markov model. IEEE Transactions on Neural Systems and Rehabilitation Engineering 13(4) (2005), 558–561
[158]. J. Katila, S. Raudaskoski, (2020). Interaction analysis as an embodied and interactive process: multimodal, co-operative, and intercorporeal ways of seeing video data as complementary professional visions. Human Studies, 43(3), 445-470.
[159]. W. Chen (2019). Knowledge-aware learning analytics for smart learning. Procedia Computer Science, 159, 1957-1965.
[160]. V. Herdel, A. Kuzminykh, A. Hildebrandt, J. R. Cauchard (2021, May). Drone in love: Emotional perception of facial expressions on flying robots. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-20).
[161]. Z. Kupper, F. Ramseyer, H. Hoffmann, S. Kalbermatten, W. Tschacher. Video-based quantification of body movement during social interaction indicates the severity of negative symptoms in patients with schizophrenia. Schizophrenia research 121(1–3) (2010), 90–100.
[162]. U. Kale. Levels of interaction and proximity: Content analysis of video-based classroom cases. The Internet and Higher Education 11(2) (2008), 119–128.
[163]. V.P. Richmond, J.C. Mccroskey, T. Mottet. Handbook of instructional communication: Rhetorical and relational perspectives. Routledge, 2015.
[164]. P. Pérez, P. Roose, Y. Cardinale, M. Dalmau, D. Masson, N. Couture (2020, November). Mobile proxemic application development for smart environments. In Proceedings of the 18th International Conference on Advances in Mobile Computing & Multimedia (pp. 94-103).

[165]. H. Kivrak, F. Cakmak, H. Kose, S. Yavuz (2021). Social navigation framework for assistive robots in human inhabited unknown environments. Engineering Science and Technology, an International Journal, 24(2), 284-298.
[166]. U. Maniscalco, P. Storniolo, A. Messina (2022). Bidirectional multi-modal signs of checking human-robot engagement and interaction. International Journal of Social Robotics, 14(5), 1295-1309.
[167]. J.S. Philpott. The relative contribution to meaning of verbal and nonverbal channels of communication: A meta-analysis. Unpublished master's thesis, University of Nebraska, Lincoln (1983).
[168]. A. Mehrabain. Some referants and measures of non-verbal behaviour, Behavioural Research Methods and Instrumentation 1 (1969), 213–217.
[169]. M. Girolami, F. Mavilia, F. Delmastro (2020). Sensing social interactions through BLE beacons and commercial mobile devices. Pervasive and Mobile Computing, 67, 101198.
[170]. R. Martínez-Maldonado, L. Yan, J. Deppeler, M. Phillips, D. Gašević (2022). Classroom analytics: Telling stories about learning spaces using sensor data. In Hybrid learning spaces (pp. 185-203). Cham: Springer International Publishing.
[171]. P.W. Miller. Nonverbal Communication. What Research Says to the Teacher., ERIC, 1988.
[172]. P.W. Miller. Body Language in the Classroom. Techniques: Connecting education and careers 80(8) (2005), 28–30.
[173]. Y. Wang, L. H. Lee, T. Braud, P. Hui (2022, July). Re-shaping Post-COVID-19 teaching and learning: A blueprint of virtual-physical blended classrooms in the metaverse era. In 2022 IEEE 42nd International Conference on Distributed Computing Systems Workshops (ICDCSW) (pp. 241-247). IEEE.
[174]. M.P. Driscoll. Psychology of learning for instruction. Needham, MA: Allyn & Bacon (2000).
[175]. S. Dhelim, H. Ning, F. Farha, L. Chen, L. Atzori, M. Daneshmand (2021). IoT-enabled social relationships meet artificial social intelligence. IEEE Internet of Things Journal, 8(24), 17817-17828.
[176]. C. Chin. Classroom interaction in science: Teacher questioning and feedback to students' responses, International Journal of Science Education 28(11) (2006), 1315–1346.
[177]. C.M. Reigeluth, J. Moore. Cognitive education and the cognitive domain, Instructional-design theories and models: A new paradigm of instructional theory 2 (1999), 51–68.
[178]. M. Teräs, J. Suoranta, H. Teräs, M. Curcher (2020). Post-Covid-19 education and education technology 'solutionism': A seller's market. Postdigital Science and Education, 2(3), 863-878.
[179]. J.A. Fredricks, P.C. Blumenfeld, A.H. Paris. School engagement: Potential of the concept, state of the evidence. Review of educational research 74(1) (2004), 59–109.
[180]. A. Nigam, R. Pasricha, T. Singh, P. Churi (2021). A systematic review on AI-based proctoring systems: Past, present and future. Education and Information Technologies, 26(5), 6421-6445.
[181]. A. Silvola, P. Näykki, A. Kaveri, H. Muukkonen. (2021). Expectations for supporting student engagement with learning analytics: An academic path perspective. Computers & Education, 168, 104192.
[182]. J. Reeve, C.-M. Tseng. Agency as a fourth aspect of students'engagement during learning activities, Contemporary Educational Psychology 36(4) (2011), 257–267.
[183]. K.J. Mach, M.C. Lemos, A.M. Meadow, C. Wyborn, N. Klenk, J.C. Arnott, ... G. Wong-Parodi (2020). Actionable knowledge and the art of engagement. Current Opinion in Environmental Sustainability, 42, 30-37.
[184]. K.R. Koedinger, J.R. Anderson, W.H. Hadley, M.A. Mark. Intelligent tutoring goes to school in the big city (1997).
[185]. L. Guo, D. Wang, F. Gu, Y. Li, Y. Wang, R. Zhou (2021). Evolution and trends in intelligent tutoring systems research: a multidisciplinary and scientometric view. Asia Pacific Education Review, 22(3), 441-461.
[186]. S.K. D'mello, A. Graesser. Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. User Modeling and User-Adapted Interaction 20(2) (2010), 147–187.
[187]. A. Behera, P. Matthew, A. Keidel, P. Vangorp, H. Fang, S. Canning (2020). Associating facial expressions and upper-body gestures with learning tasks for enhancing intelligent tutoring systems. International Journal of Artificial Intelligence in Education, 30, 236-270.
[188]. E. Joseph. Engagement tracing: using response times to model student disengagement. Artificial intelligence in education: Supporting learning through intelligent and socially informed technology 125 (2005), 88.

[189]. S. Li, S. P. Lajoie, J. Zheng, H. Wu, H. Cheng (2021). Automated detection of cognitive engagement to inform the art of staying engaged in problem-solving. Computers & Education, 163, 104114.
[190]. M. Chaouachi, C. Pierre, I. Jraidi, C. Frasson. Affect and mental engagement: Towards adaptability for intelligent. In Twenty-Third International FLAIRS Conference, 2010.
[191]. B.S. Goldberg, R.A. Sottilare, K.W. Brawner, H.K. Holden. Predicting learner engagement during well-defined and ill-defined computer-based intercultural interactions. In International Conference on Affective Computing and Intelligent Interaction, Springer, 2011, pp. 538–547.
[192]. X. Xiao, J. Wang. Understanding and detecting divided attention in mobile MOOC learning. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 2017, pp. 2411–2415.
[193]. F. Alqahtani, S. Katsigiannis, N. Ramzan (2020). Using wearable physiological sensors for affect-aware intelligent tutoring systems. IEEE Sensors Journal, 21(3), 3366-3378.
[194]. J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, J.R. Movellan. The faces of engagement: Automatic recognition of student engagementfrom facial expressions. IEEE Transactions on Affective Computing 5(1) (2014), 86–98.
[195]. C.R. Beal, R. Walles, I. Arroyo, B.P. Woolf. On-line tutoring for math achievement testing: A controlled evaluation. Journal of Interactive Online Learning 6(1) (2007), 43–55.
[196]. A. Kaur, A. Mustafa, L. Mehta, A. Dhall. Prediction and localization of student engagement in the wild. In 2018 Digital Image Computing: Techniques and Applications (DICTA), IEEE, 2018, pp. 1–8.
[197]. T. Baltrušaitis, P. Robinson, L.-P. Morency. Openface: an open-source facial behavior analysis toolkit. In 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2016, pp. 1–10.
[198]. B. Zhu, X. Lan, X. Guo, K. E. Barner, C. Boncelet. (2020, October). Multi-rate attention based GRU model for engagement prediction. In Proceedings of the 2020 International Conference on Multimodal Interaction (pp. 841-848).
[199]. Y. Y. Li, Y. P. Hung (2019, September). Feature fusion of face and body for engagement intensity detection. In 2019 IEEE international conference on image processing (ICIP) (pp. 3312-3316). IEEE.
[200]. V. Thong Huynh, S. H. Kim, G. S. Lee, H. J. Yang (2019, October). Engagement intensity prediction with Facial behavior features. In 2019 International Conference on Multimodal Interaction (pp. 567-571).
[201]. P. Demochkina, A. Savchenko (2022, September). Efficient Algorithms for Video-Based Engagement Prediction for a MOOC Course. In 2022 International Russian Automation Conference (RusAutoCon) (pp. 672-676). IEEE.
[202]. S. Wu (2021). Simulation of classroom student behavior recognition based on PSO-kNN algorithm and emotional image processing. Journal of Intelligent & Fuzzy Systems, 40(4), 7273-7283.
[203]. S. Chakraborty, R. Mondal, P. K. Singh, R. Sarkar, D. Bhattacharjee (2021). Transfer learning with fine tuning for human action recognition from still images. Multimedia Tools and Applications, 80, 20547-20578.
[204]. A. Nadeem, A. Jalal, K. Kim (2021). Automatic human posture estimation for sport activity recognition with robust body parts detection and entropy Markov model. Multimedia Tools and Applications, 80, 21465-21498.
[205]. I. Akhter, A. Jalal, K. Kim (2021, January). Pose estimation and detection for event recognition using Sense-Aware features and Adaboost classifier. In 2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST) (pp. 500-505). IEEE.
[206]. N. Irvine, C. Nugent, S. Zhang, H. Wang, W. W. Ng (2019). Neural network ensembles for sensor-based human activity recognition within smart environments. Sensors, 20(1), 216.
[207]. Y. Y. Ghadi, I. Akhter, S. A. Alsuhibany, T. al Shloul, A. Jalal, K. Kim (2022). Multiple events detection using context-intelligence features. INTELLIGENT AUTOMATION AND SOFT COMPUTING, 34(3).
[208]. G. Mohmed, A. Lotfi, A. Pourabdollah (2020, June). Employing a deep convolutional neural network for human activity recognition based on binary ambient sensor data. In Proceedings of the 13th ACM international conference on pervasive technologies related to assistive environments (pp. 1-7).
[209]. Z. Ahmad, N. M. Khan (2019, September). Multidomain multimodal fusion for human action recognition using inertial sensors. In 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM) (pp. 429-434). IEEE.

[210]. M. Wang, Z. Yan, T. Wang, P. Cai, S. Gao, Y. Zeng, ..., X. Chen (2020). Gesture recognition using a bioinspired learning architecture that integrates visual data with somatosensory data from stretchable sensors. Nature Electronics, 3(9), 563-570.
[211]. M. Dewan, M. Murshed, F. Lin (2019). Engagement detection in online learning: a review. Smart Learning Environments, 6(1), 1-20.
[212]. Y. Du, R. G. Crespo, O. S. Martínez (2023). Human emotion recognition for enhanced performance evaluation in e-learning. Progress in Artificial Intelligence, 12(2), 199-211.
[213]. L. Wang, K. J. Yoon (2021). Knowledge distillation and student-teacher learning for visual intelligence: A review and new outlooks. IEEE transactions on pattern analysis and machine intelligence, 44(6), 3048-3068.
[214]. E. Baran, D. Alzoubi (2020). Human-centered design as a frame for transition to remote teaching during the COVID-19 pandemic. Journal of Technology and Teacher Education, 28(2), 365-372.
[215]. E. Fonseca, X. Favory, J. Pons, F. Font, X. Serra (2021). Fsd50k: an open dataset of human-labeled sound events. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 30, 829-852.
[216]. R. Zheng, F. Jiang, R. Shen (2020, May). Intelligent student behavior analysis system for real classrooms. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 9244-9248). IEEE.
[217]. C. Pabba, P. Kumar (2022). An intelligent system for monitoring students' engagement in large classroom teaching through facial expression recognition. Expert Systems, 39(1), e12839.
[218]. S. E. Ovur, H. Su, W. Qi, E. De Momi, G. Ferrigno (2021). Novel adaptive sensor fusion methodology for hand pose estimation with multileap motion. IEEE Transactions on Instrumentation and Measurement, 70, 1-8.
[219]. J. Xu, Z. Yu, B. Ni, J. Yang, X. Yang, W. Zhang (2020). Deep kinematics analysis for monocular 3d human pose estimation. In Proceedings of the IEEE/CVF Conference on computer vision and Pattern recognition (pp. 899-908).
[220]. P. Goldberg, Ö. Sümer, K. Stürmer, W. Wagner, R. Göllner, P. Gerjets, ..., U. Trautwein (2021). Attentive or not? Toward a machine learning approach to assessing students’ visible engagement in classroom instruction. Educational Psychology Review, 33, 27-49.
[221]. T. Shou, C. Borchers, S. Karumbaiah, V. Aleven (2023). Optimizing parameters for accurate position data mining in diverse classrooms layouts.
[222]. A. Bigalke, L. Hansen, J. Diesel, C. Hennigs, P. Rostalski, M. P. Heinrich (2023). Anatomy-guided domain adaptation for 3D in-bed human pose estimation. Medical Image Analysis, 102887.
[223]. C. Luo, J. Zhang, J. Yu, C. W. Chen, S. Wang (2019). Real-time head pose estimation and face modeling from a depth image. IEEE Transactions on Multimedia, 21(10), 2473-2481.
[224]. G. D. Abowd, C. G. Atkeson, A. Feinstein, C. Hmelo, R. Kooper, S. Long, N. Sawhney, M. Tani, Teaching and learning as multimedia authoring: the classroom 2000 project. In Proceedings of the fourth ACM international conference on Multimedia, 1997, pp. 187–198.
[225]. L. Zhang, S. Lin. Research on the Design and Application of Intelligence Classroom Teaching Model with Rain Classroom Digital Support. In International Conference on Modern Educational Technology and Innovation and Entrepreneurship (ICMETIE 2020), Atlantis Press, 2020, pp. 368–373.
[226]. M. A. Mady, S. Baadel. Technology-Enabled Learning (TEL): YouTube as a Ubiquitous Learning Aid. Journal of Information & Knowledge Management 19(01) (2020), 2040007.
[227]. J. C. Augusto. Ambient intelligence: Opportunities and consequences of its use in smart classrooms. Innovation in Teaching and Learning in Information and Computer Sciences 8(2) (2009), 53–63.
[228]. I. Abdellatif. Towards A Novel Approach for Designing Smart Classrooms. In 2019 IEEE 2nd International Conference on Information and Computer Technologies (ICICT), IEEE, 2019, pp. 280–284.
[229]. S. Jaiswal, A. Parmar, H. Singh, G. Rathee. Smart Classroom Automation (2018).
[230]. G. Basilaia, D. Kvavadze. Transition to online education in schools during a SARS-CoV-2 coronavirus (COVID-19) pandemic in Georgia. Pedagogical Research 5(4) (2020), 1–9.
[231]. A. Magnani. Human Action Recognition and Monitoring in Ambient Assisted Living Environments. PhD thesis, alma, 2020.
[232]. G. Cebrian, R. Palau, J. Mogas. The Smart Classroom as a means to the development of ESD methodologies. Sustainability 12(7) (2020), 3010.
[233]. L. Chen, J. Hoey, C. D. Nugent, D.J. Cook, Z. Yu. Sensor-based activity recognition. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 42(6) (2012), 790–808.

[234]. B. Logan, J. Healey, M. Philipose, E.M. Tapia, S. Intille. A long-term evaluation of sensing modalities for activity recognition. In International conference on Ubiquitous computing, Springer, 2007, pp. 483–500.

[235]. E.M. Tapia, S.S. Intille, W. Haskell, K. Larson, J. Wright, A. King, R. Friedman. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. In 2007 11th IEEE international symposium on wearable computers, IEEE, 2007, pp. 37–40.

[236]. M. Stikic, T. Huynh, K. Van Laerhoven, B. Schiele. ADL recognition based on the combination of RFID and accelerometer sensing. In 2008 second international conference on pervasive computing technologies for healthcare, IEEE, 2008, pp. 258–263.

[237]. N. Roy, A. Misra, D. Cook. Ambient and smartphone sensor assisted ADL recognition in multi-inhabitant smart environments. Journal of ambient intelligence and humanized computing 7(1) (2016), 1–19.

[238]. T. Diethe, N. Twomey, M. Kull, P. Flach, I. Craddock, Probabilistic sensor fusion for ambient assisted living. arXiv preprint arXiv:1702.01209 (2017).

[239]. A. Hermanis, R. Cacurs, K. Nesenbergs, M. Greitans, E. Syundyukov, L. Selavo. Wearable sensor system for human biomechanics monitoring. In Proceedings of the 2016 International Conference on Embedded Wireless Systems and Networks, 2016, pp. 247–248.

[240]. S. Jung, S. Hong, J. Kim, S. Lee, T. Hyeon, M. Lee, D.-H. Kim. Wearable fall detector using integrated sensors and energy devices. Scientific reports 5 (2015), 17081.

[241]. T.T. Um, V. Babakeshizadeh, D. Kuli´c. Exercise motion classification from large-scale wearable sensor data using convolutional neural networks. In 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2017, pp. 2385–2390.

[242]. Z. Wang, S.Y. He, Y. Leung. Applying mobile phone data to travel behaviour research: A literature review. Travel Behaviour and Society 11 (2018), 141–155.

[243]. E. Dimitriadou, A. Lanitis. (2023). A critical evaluation, challenges, and future perspectives of using artificial intelligence and emerging technologies in smart classrooms. Smart Learning Environments, 10(1), 1-26.

[244]. M. Kwet, P. Prinsloo. (2020). The ‘smart’classroom: a new frontier in the age of the smart university. Teaching in Higher Education, 25(4), 510-526.

[245]. L. Chen, C. D. Nugent. (2019). Human activity recognition and behavior analysis. Springer International Publishing.

[246]. J. Candamo, M. Shreve, D. B. Goldgof, D. B. Sapper, R. Kasturi. (2009). Understanding transit scenes: A survey on human behavior-recognition algorithms. IEEE transactions on intelligent transportation systems, 11(1), 206-224.

Информация об авторах / Information about authors

Мария Луиза КОРДОБА-ТЛАКСКАЛЬТЕКО – лектор факультета статистики и информатики Университета Веракруса. Сфера научных интересов: формальные языки, совместные вычисления, искусственный интеллект.

María Luisa CÓRDOBA-TLAXCALTECO – Lecturer of the Faculty of Statistics and Informatics of the University of Veracruz in Mexico. Research interests: formal languages, collaborative computing, artificial intelligence.

Эдгард БЕНИТЕС-ГЕРРЕРО – профессор факультета статистики и информатики Университета Веракруса. Сфера научных интересов: человеко-машинное взаимодействие, искусственный интеллект, системы управления данными, совместные вычисления.

Edgard BENÍTEZ-GUERRERO – Professor of the Faculty of Statistics and Informatics of the University of Veracruz in Mexico. Research interests: human-Computer Interaction, artificial intelligence, data management systems, collaborative computing.