Sensors and Materials, Vol. 36, No. # (2023) 1–23
MYU Tokyo

S & M ####

Visual Human Behavior Sensing and Understanding for Autism Spectrum Disorder Treatment: A Review

Xuna Wang,[1] Hongwei Gao,[1*] Yutong Zhang,[1] Yueqiu Jiang,[1] and Jiahui Yu[2,3**]

[1]School of Automation and Electrical Engineering, Shenyang Ligong University, Shenyang City, Liaoning 110159, China
[2]Department of Biomedical Engineering, Zhejiang University, Hangzhou 310027, China
[3]Innovation Center for Smart Medical Technologies & Devices, Binjiang Institute of Zhejiang University, Hangzhou 310053, China

(Received April 25, 2023; accepted December 18, 2023)

Keywords: autism spectrum disorder, computer vision, human action understanding, visual sensing

With the increasing incidence of autism spectrum disorders (ASDs), the physical and mental status of patients and the economic burden on their families have become a major concern. Individuals with ASD exhibit diverse motor characteristics. The combination of contactless visual sensors and computer vision (CV) technology allows the noncontact and long-term monitoring of these characteristics to extract valuable quantitative information. Therefore, in this paper we systematically review CV technology using visual sensors to obtain motion information from individuals with ASDs with the aim of exploring the application of noncontact perception systems in the diagnosis and treatment of autism. (1) A systematic review of publications indexed on Web of Science, PubMed, and Engineering Village and studies published from January 2015 to March 2023 was conducted. (2) Different publicly available datasets were reviewed to accelerate related research. (3) We summarized the above research results in tables and analyzed the research status, open challenges, and future perspectives. The results of this review show that the use of visual sensors to capture human movement information has wide application value in the diagnosis and treatment of autism.

1. Introduction

The incidence of autism spectrum disorders (ASDs) is increasing yearly worldwide, which has rapidly developed into a global public health crisis. The social function of children with a long-term chronic course of the disease has varying degrees of impairment. However, the early intervention rate is low owing to the severe lag in the early screening, diagnosis, and treatment of autism.[(1)] The long early intervention period for children with autism and the high cost of diagnosis and treatment have a severe economic burden on many families. In addition, the serious imbalance between professional caregivers and patients has also brought tremendous work burdens and pressure to practitioners. These limitations and the increasing prevalence of

*Corresponding author: e-mail: ghw1978@sohu.com
**Corresponding author: e-mail: jiahui.yu@zju.edu.cn

ISSN 0914-4935
https://myukk.org/

ASDs require the development of more automated and accurate sensing systems to reduce rehabilitation costs and evaluation time.

Because autistic people have diverse motor characteristics, observing and analyzing children's natural movements can help in the early detection of risk. ASD diagnostic features fall into two main categories: (1) persistent deficiencies in social communication and social interaction in multiple settings,(2) i.e., patients lack direction, response, and sharing in social communication, which may be accompanied by poor integration of body postures,(3) and (2) restricted and repetitive behavior, interest, or activity patterns.(2) Some prominent signs are hand slapping, body swaying, rotation, repetitive jumping, and finger flicking.(4) Associated features of ASDs are divided into three categories: motor deficits, disruptive/challenging behaviors, and catatonic motor behavior.(2)

In medical environments, contactless visual sensors primarily include RGB cameras and 3D depth cameras, which offer a cost-effective and straightforward approach to capturing children's behavioral patterns in a noninvasive and continuous manner over time. In contrast, wearable devices can restrict the movements of children with autism and even trigger anxiety and self-stimulatory behaviors.(5) Furthermore, wearable devices cannot detect external assistance or interactions with the physical environment.(6) In recent years, with the widespread application of computer vision (CV) techniques in various fields,(7,8) there has been a noticeable increase in the amount of research utilizing patients' motion information for automatic quantitative analysis. The combination of visual sensors and CV technology enables the noncontact and long-term monitoring of autism risk signals while providing objective quantitative evidence.(6)

Recent reviews and articles have demonstrated the utility of such tools in individuals with ASD. De Belen *et al.*(9) and Sadek *et al.* (5,10) introduced CV to capture and quantify various information in ASD diagnosis. To automate the assessment of motor impairments in ASD, Thomas *et al.*(11) systematically explored computational methods, including devices such as accelerometers for contact-based measurements. Unlike previous work, we explore the feasibility and scope of an autism intelligence system that uses motion information only obtained by contactless visual sensors, thereby increasing the complexity of the review process. Furthermore, this study encompasses the fields of technology, artificial intelligence (AI) algorithms, and autism medicine, further complicating the analysis process. Our main contributions are as follows: (1) We design a retrieval method and inclusion criteria for research focused on vision-based behavior perception and understanding in patients with ASD. (2) We introduce each qualified paper according to the category and summarize the information in tables. (3) We introduce publicly available ASD datasets that include visual motion information. (4) We analyze the research status and summarize the open challenges and future perspectives.

2. Data, Materials, and Methods

The research scope of this project is limited to the acquisition of human motion data using noncontact devices, where the human motion information does not include eye-tracking and facial motion capture. In addition, we provide a summary of the relevant public datasets collected during the survey. The co-development and two-way communication between ASD medicine and AI technology has led to valuable results and mature technologies. Considering the timeliness of the technology, only relevant literature from January 2015 to March 2023 has been systematically reviewed.

Information sources and retrieval strategies. To guarantee the quality of reference papers, we obtained relevant literature from three traditional channels: Web of Science, PubMed, and Engineering Village. These three channels belong to the secondary literature query system, which has specific quality requirements for the collected documents. The search strategies are as follows:

(1) Theme: 'autis*' AND ('movement' OR 'behavio*' OR 'acti*') AND ('visual' OR 'vision' OR 'imag*' OR 'video') AND ('automatic' OR 'comput*' OR 'engineering')
(2) Theme: 'autis*' AND ('movement analysis' OR 'behavio* imaging' OR 'behavio* analysis' OR 'acti* recognition')
(3) Theme: 'autis*' AND ('movement' OR 'behavio*' OR 'acti*'); Research direction: Computer Science

Inclusion criteria. The title, abstract, and method of each article were scanned for relevance. Specific criteria are as follows: (1) The research object contains autistic patients. (2) Motion information was obtained only from contactless visual sensors. (3) Motion information includes complete or partial body motions, excluding gaze, facial expressions, and sleep. (4) The processing and analysis of information are automated. (5) Results in the form of patents, reviews, meta-analyses, keynotes, narratives, or editorials are excluded. (6) Papers are written in English. (7) The complete literature can be searched.

Data entry. Through the above search process, 63 eligible studies were included. Among them, 42 studies were related to diagnosis and 21 were related to intervention. Where possible, we extracted the following information from each study into an Excel spreadsheet: (1) the intervention or diagnosis method used, (2) the autism characteristics studied, (3) the reference number of the paper, (4) the CV task for processing motion information, (5) the specific CV method for obtaining motion information, and (6) the sensor configuration for collecting motion information. In addition, 13 autism datasets with links to resources were found.

3. Results

3.1 Related work

In recent years, with the development of CV technology and patient demand for contactless diagnosis and treatment, the use of CV to analyze the motion information of autistic patients has increased. In this review, we provide ample evidence of the effectiveness of such techniques in (1) identifying and quantifying behavioral markers for the diagnosis and assessment of ASDs and (2) constructing unconstrained therapies or adjunct tools.

3.1.1 Autism diagnosis

CV-based systems provide a low-cost and noninvasive diagnostic method and can reduce the errors associated with human factors in decision-making.[5] According to the characteristics of the autism category and the corresponding research situation, the relevant works are divided into social behavior disorder, atypical behavior (not limited to social influence), motor deficits, and abnormal body posture. Studies related to autism diagnosis are summarized in Tables 1–3, and each quantified information is described in Sect. 2.

3.1.1.1 Social communication and social interaction

By quantifying the motion information generated during social interactions, the extent of impaired interpersonal behavior coordination in subjects can be objectively assessed, aiding in

Table 1
Related works on social communication and social interaction.

ASD characteristic	Ref.	CV task	Specific methods	Sensor placement: number of sensors
Pointing behavior	12	3D pose estimation	Microsoft Kinect SDK	Kinect: 1
	13	Quantification of hand movement	YOLOv3, ResNet-18, OpenPose	RGB camera: 1
Response to name	14	Quantification of head movement	PCA, face detection/alignment, head pose feature extraction	RGB camera: 1
	15	Quantification of head movement	CVA	Tablet camera: 1
	16	Quantification of head movement	IntraFace	Tablet camera: 1
	17	Quantification of head movement	IntraFace, tracking facial landmarks	Tablet camera: 1
	18	Gesture recognition	VGG16/SSD	RGB camera: 2; Kinect: 1
	19	Quantification of head movement	YOLOv3, HRNet, OSNet, OpenFace	RGB camera: 4
Response to instructions	20	Quantification of head movement	SSD	Logitech BRIO: 2; Kinect: 1
	21	Action recognition	OstAD, YOLOv5, I3D	RGB camera: 3
Atypical attention	22	Quantification of head movement	CVA	Tablet camera: 1
Movement synchrony	23	Gesture assessment, action recognition	MMSN	No relevant introduction
Atypical behaviors	3	Quantification of motor patterns	MEA	RGB camera: 1
	25	Motion feature extraction	OpenPose, Gaussian mixture model, bidirectional long short term memory neural network	RGB camera: 2
	26	Action recognition	OpenPose, VGG16-LSTM	No relevant introduction
	27	Quantification of head movement	OpenPose, SVM	Wireless Ezviz CS-C2C-1B2WFR camera: 4
	39	Quantification of motor patterns	MEA, NeuroMiner, SVM with linear kernel	RGB camera: 1

Table 2
Related works on atypical behaviors (not limited to social).

ASD characteristic	Ref.	CV task	Specific methods	Sensor placement: number of sensors
Disruptive behaviors	35	Action recognition	OpenPose, time-distributed CNN, LSTM	Existing dataset (SSBD)
Stereotyped behavior	28	Action recognition	Kinect for Windows SDK, $P Point-Cloud Recognizer	Kinect: 1
	31	Pose estimation	2D Mask R-CNN	Data collected from NODA program
	32	Action recognition	OpenPose	Data collected from YouTube
	33	Action recognition	O-GAD	No relevant introduction
	34	Action recognition	AlphaPose, 3DCNN /ConvLSTM	Collected data, equipment not fixed
	38	Action recognition	OpenPose, LSTM	5 million pixels Hikvision remote camera: 4
	39	Action recognition	CSRT, HRNet, RGBPose-SlowFast	Existing dataset (SSBD, Autism dataset)
	40	Action recognition	I3D/TSN, weak supervision	Existing dataset (HMDB51, SSBD)
	41	Action recognition	CNN, transfer learning	Webcam
	42	Quantification of motor patterns	OpenNI/NITE framework	Kinect: 1

Table 3
Related works on motor deficits and abnormal body posture.

ASD characteristic	Ref.	CV task	Specific methods	Sensor placement: number of sensors
Whole body posture	44	Quantification of motor patterns	MOVIDEA	RGB camera: 1
	45	Classification	Linear discriminant analysis/logistic regression/multilayered perceptron/log-linearized Gaussian mixture network	RGB-D camera: 1
	46	Quantification of motor patterns	Mask R-CNN, OpenPose, Spearman’s correlation coefficient	No relevant introduction
	47	Classification	OpenPose, SVM	RGB-D camera: 1
	48	Quantification of motor patterns	Motognosis	RGB camera: 1
	49	2D pose estimation	Processing software	RGB camera: 1
Head posture	50	Quantification of head movement	Zface	RGB camera: 1
	51	Quantification of head movement	CVA	Tablet camera: 1
	52	Quantification of head movement	Dlib-ml, OpenFace, head pose feature extraction	Tablet camera: 1
	53	Quantification of head movement	OpenFace	RGB camera: 1
Hand posture	55	Classification	LSTM model	RGB camera: 1
	56	Classification	InceptionV3/ResNet-50, 2-layer LSTM	Existing dataset (unavailable)
	57	Classification	Spatial attention bilinear pooling, LSTM	Existing dataset (ASDD)
	58	Gesture assessment	OpenCV, cosine similarity formula	Mobile camera: 1

diagnostic analysis. Three standard tests assess social communication and interaction in people with autism.

The first is the Expressing Needs with Pointing Test, in which hand pointing is an essential source of reaction information. Wang *et al.*(12) proposed a detailed protocol to describe this clinical task. In this work, mutual gaze and gesture were the primary basis for judging the children's performance. Qin *et al.*(13) further developed the evaluation method, in which features such as hand position, gesture, and pointing direction were detected. The accuracy of this automated evaluation system was 17/19, indicating that the directive behaviors expressing needs are accurately assessed.

The second is the Response to Name Test, for which gaze estimation, head posture, and shoulder posture are essential sources of response information. Liu *et al.*(14) proposed a dataset and an automated prediction system that considered the response speed, eye contact duration, and head direction to output responsiveness scores. Campbell *et al.*,(15) Perochon *et al.*,(16) and Hashemi *et al.*(17) employed video stimuli to capture children's attention. The latter two studies recorded the naming response and encoded the response latency. It was found that children with autism exhibited a lower response frequency and a longer response time. Wang *et al.*(18) and Song *et al.*(19) used toys to attract children's attention. Later work also considered the shoulder rotation angle, and the experiment achieved a high classification accuracy of 93.3%.

The third is the Response to Instructions Test, for which gaze estimation and response action detection are the primary sources of response information. Liu *et al.*(20) proposed a protocol in which a clinician asks a child to hand them a toy for interactive play. Shi *et al.*(21) designed the Ost-AD network for the protocol proposed in Ref. 20. This network used temporal attention branches to aggregate contextual features and spatial attention branches to generate local frame-level features of children. Their model achieved more than 70% classification accuracy but still needs improvement.

In addition, studies have shown that, compared with nonsocial stimulation, ASD patients pay less attention to social interaction. Bovery *et al.*(22) used two sides of a screen to display social and nonsocial stimuli. Then they detected the subject's direction of attention by analyzing head and iris positions. Difficulties in social interaction with individuals with ASD are also reflected in the dynamic temporal connection between the motions of interacting people. Li *et al.*(23) proposed a network that automatically assessed a child's movement synchronization with a therapist. In this network, inflated 3D convolutional neural networks (CNNs) were used for feature extraction, and three output heads were used for specific tasks: motion quality assessment, motion synchronization prediction, and intervention identification. In addition, the authors applied label distribution learning to mitigate the artificial bias in motion synchronization estimation. In this work, they produced an outcome comparable to those of standard methods at a much reduced cost.

Some studies used subjects' behavior during social interaction to determine whether they have ASDs. Georgescu *et al.*(24) used a support vector machine (SVM) with a linear kernel to classify high-functioning ASD adults and typical developing (TD) adults. The parameters included intrapersonal synchrony between the head and body, which was quantified using motion energy analysis and the open-source machine learning tool NeuroMiner. The accuracy of this method was 75.9%. Lin *et al.*[25] designed a multimodal (speech acoustics and body gestures) interlocutor-modulated attention network architecture to differentiate between the three ASD subgroups. The motion part was based on gestural features derived from the tracked body joints of each frame. The network achieved an unweighted average recall rate of 66.8%, which could be improved. Kojovic *et al.*[26] distinguished children with ASD from children with TD using skeletal information generated from videos of social interaction. They used a CNN combined with long short term memory (LSTM) to classify the action. The input of this architecture was an image with a deleted background, not the original key-point coordinates. The accuracy of this model was 80.9%. Tang *et al.*[27] analyzed children's head movements, facial expressions, and vocal features under different attitudes toward their mothers. The accuracy of the SVM classifier was more than 90%.

3.1.1.2 Atypical behaviors (not limited to social)

Stereotyped movements are semi-voluntary repetitive movements, a prominent clinical feature of ASDs. The head, wrist, elbow, and shoulder joint are the key points in their investigation. Maha *et al.*[28] automatically detected atypical motions using point clouds. This method only considers spatial information, but temporal information is also an important reference factor in behavior recognition tasks.[29,30] On this basis, Kathan *et al.*[31] and Cook *et al.*[32] calculated movement information. In contrast to the manually designed time features, Tian *et al.*[33] used a 3D CNN to generate shared time feature maps from videos automatically. They proposed a new time pyramid network to mine features at different semantic levels. These features were used for tasks of varying granularity: short-term ASD-related action detection and long-term repetitive behavior recognition. Negin *et al.*[34] used LSTM to learn the temporal evolution of skeleton sequences and emphasized the importance of detecting children's behavior in an uncontrolled environment. Compared with human-designed features, auto-encoded features have better generalization. In addition to analyzing trunk and limb movement information, some studies evaluated patients using head or hand movement information. Head banging is one of the stimming behaviors of autistic patients, which harms the patients themselves and needs timely outside intervention. Accordingly, Washington *et al.*[35] designed a skeletal CNN-LSTM network and achieved a mean F1-score of 90.77% for recognizing head banging behavior. Hand motion complexity is associated with limiting repetitive and stereotyped behavior; diversity is associated with behaviors critical to independent living.[36] Zhang *et al.*[37] proposed a strategy for applying gesture recognition to skeletal datasets to explore subfeatures. They compressed the middle layer output feature map using an hourglass-structured convolutional network, which was then mapped to classified subfeatures.

For uncontrolled environments, some studies considered the behavior detection of multiple people. Zhang *et al.*[38] matched the distance between current and previous skeletons to track various children with ASDs in the same scene. Pandian *et al.*[39] initialized the tracker using manually annotated initial frame bounding boxes. In this work, they fed raw video signals and a skeletal joint thermal map into the proposed depth network. Lack of data is another common problem with deep neural networks. Accordingly, Pandey *et al.*[40] proposed a technique called guided weak supervision. In this work, they utilized optical flow frames for category matching because the optical flow transform covers most nonmotion-related information and exaggerates the motion information.

As a diagnostic feature of autism, repetitive behavior has many reference values in ASD medicine. First, the assessment of repetitive behaviors plays a crucial role in the prescription of medication dosages. To address the issue of needing continuous monitoring with abnormal behavior checklist, Prabha *et al.*(41) used deep convolutional networks and transfer learning to assess the repetitive behaviors of children with ASD. The method was validated by drug temperature regulation in children with autism. In addition, stereotyped behavior may be associated with lower mental health. Camada *et al.*(42) proposed the use of machine learning algorithms to identify repetitive behaviors and determine appropriate activation levels. Adaptive neural fuzzy technology based on the fuzzy C-means algorithm was used to determine the activation level of stereotyped behavior.

3.1.1.3 Motor deficits and abnormal body posture

Motor difficulties in individuals with ASD can be quantified and treated. It is suggested that efforts aimed at detecting and intervening in motor function may also positively impact social communication.(43) Motor deficits are potential early markers and predictors of ASD diagnosis. Accordingly, Caruso *et al.*(44) and Hirokazu *et al.*(45) evaluated the free movement of infants at high risk of ASD. They revealed that the signs of ASD risk could be detected as early as four months after birth by focusing on the infant's spontaneous bodily movements. In addition, Zhao *et al.*(3) used image differencing technique to extract motion time series from video records. Then they performed spectral analysis to quantify the average power of motion and the fractal scaling of movement. Jin *et al.*(46) quantified pixel distance and instantaneous pixel velocity as motion features of ASD children. Mariano *et al.*(47) studied changes in body movement tracked by depth sensor cameras under visual, auditory, and olfactory stimuli in a multimodal virtual reality (VR) experience. The authors characterized the level of movement by calculating the average displacement of the joints, achieving an accuracy of 89.36%. Owing to the prioritization of early diagnosis and treatment of autism, there is little information about motor function in adult ASDs. On this basis, Cho *et al.*(48) proposed motion tests for adult ASDs and analyzed the depth data to obtain the performance of standing, walking, and repetitive movements.

Postural control is a motor ability developed in childhood, which is reflected in maintaining stable head and body posture without excessive rocking. Children with autism often have a high deviation angle characteristic. Khan *et al.*(49) used the humerus as the baseline and measured the angle of the arm moving outwards in a regular standing position. Moreover, children with ASD may use head movements to regulate their perception of social situations. Martin *et al.*,(50) Dawson *et al.*,(51) and Babu *et al.*(52) obtained head motion data by calculating facial feature points, while Zhao *et al.*(53) focused on temporal change descriptors extracted from head motion feature sequences. The results showed that the ASD group had significantly higher levels of pitch (head point), yaw (head turn), roll (head roll), head rotation range, and average rotation per minute in their head movement and that the degree of head motion was not positively correlated with the interlocutor's visual gaze.

Even in the initial stages of gestures, motor behavior is embedded with information about its intention to perform.(54) Some researchers have investigated how ASDs affect intentionality in the initial stages of gestures. Andrea *et al.*(55) and Pandya *et al.*(56) used pretrained GoogleNet with LSTM to classify ASDs. The former proposed a new dataset including matched-IQ ASD and TD children. Two groups of children were required to grab a bottle and perform four different follow-up actions: placing, pouring, passing to pour, and passing to place. Using the same dataset, Sun *et al.*(57) proposed bilinear pooling of spatial attention to enhance spatial information extraction without significantly increasing the number of parameters, which can dynamically and effectively focus on more discriminative regions. The average accuracy of this work reached 82.56%. On the other hand, rare signs of neurological disorders in the hands of autistic people, such as small gaps, are seen as the first signs of autism. Shushma *et al.*(58) used OpenCV for 2D posture detection and then the cosine similarity formula to distinguish the gap between fingers. However, the effectiveness of this work needs to be further verified.

3.1.2 Autism intervention

In this part, we introduce relevant work from two perspectives: (1) noncontact therapy methods popular in recent years, namely, game-assisted therapy, music-assisted therapy, and robot-assisted therapy, and (2) the participation and psychological state of the patient in the treatment process. The intervention studies on autism are summarized in Tables 4 and 5, and each quantified information is described in Sect. 2.

3.1.2.1 Rehabilitation games

Rehabilitation games are effective in improving children's physical and cognitive skills while giving them a lighthearted experience and are therefore seen as an effective treatment for children diagnosed with ASD.(59) Motion-sensing games are the main form of rehabilitation games. Kinect usually outputs 3D coordinates for subsequent interaction detection and action recognition. Piana *et al.*(60) developed a game prototype in which children were asked to guess the emotion or express the sentiment with postural body gestures. They used the EyesWeb and PADDLE machine learning libraries to learn hectographs and adaptive descriptors from 3D motion data and finally identified body emotions by a linear SVM. Ma *et al.*(61) and Wang *et al.*(62) used somatosensory games to test motor function coordination. In the games, touch detection and motion recognition of the body and props on the screen were performed by calculating 3D coordinate points obtained by Kinect.

Augmented reality (AR)/virtual reality (VR) can bring teaching situations to different places and contribute to innovation of the teaching paradigm. Ahlers *et al.*(63) proposed an AR computer interaction system based on speech and gesture interaction that aimed to improve children's cognitive abilities and reduce the burden on teachers. Unlike VR, where participants have to wear special glasses, the immersive 3D VR environment allows participants to literally walk into the training environment. This helps children with ASD concentrate and feel a sense of stability. Accordingly, Tsai *et al.*(64) used a third-person perspective role-playing game to teach social skills and help deepen understanding of basic emotions.

Table 4
Related works on autism intervention.

Intervention method	ASD characteristic	Ref.	CV task	Specific methods	Sensor placement: number of sensors
Rehabilitation games	Comprehensive abilities	60	Recogntion of emotions	EyesWeb, PADDLE, SVM	Kinect: 1
		61	Pose estimation	Microsoft Kinect SDK	Kinect:1
		63	3D pose estimation	Microsoft Kinect SDK	RGB camera: 1; Kinect: 1
	Motor deficits	62	Quantification of motor patterns	Microsoft Kinect SDK	RGB camera: 1; Kinect: 1
	Emotion recognition	64	3D pose estimation	Microsoft Kinect SDK	RGB camera: 1; Kinect: 2
Music therapy	Motor deficits	67	3D pose estimation	OptiTrack	Kinect: 1
		68	Quantification of motor patterns	Gesture tracking algorithm, gesture parameters detection algorithm	RGB camera: 1
	Comprehensive abilities	69	Quantification of motor patterns	Microsoft Kinect SDK, MEA	Kinect: 1; GoPro camera: 1
Robot-assisted therapy	Motor deficits	73	Estimation of rhythmic motion timing	OpenPose, RNN, FFT	USB monocular camera: 1
	Posture imitation	76	Action recognition	Rule-based finite state machine	Kinect: 1
		77	Action recognition	Sensory-motor association paradigm	Robot visual sensor
		78	Quantification of motor patterns	Microsoft Kinect SDK, HMM, GMM	RGB camera: 2; Kinect: 1
	Atypical behaviors	83	Quantification of head movement	Original conditional local neural field	Tablet camera: 1
		84	Action recognition	3D MTG	Kinect: 1
		85	Action recognition	Nuitrack SDK, CNN	Intel Real Sense 3D sensor: 1

Table 5
Related works on engagement.

Intervention method	Ref.	CV task	Specific methods	Sensor placement: number of sensors
Rehabilitation training	87	3D pose estimation	Microsoft Kinect SDK, SVM	Kinect: 1 HD camera: 1
Neurofeedback therapy	88	Quantification of motor patterns	SSD, local binary pattern, active appearance model, PnP, image gradient random forest classifier	Webcam
Robot-assisted therapy	90	Quantification of motor patterns	OpenPose	Monitor camera: 1
	91	2D pose estimation, hand pose estimation	OpenPose, E4 ACC	Robot’s visual sensor
	92	Action recognition	Neural networks, transfer learning	Existing dataset (unavailable)
	93	Quantification of motor patterns	OpenFace	Robot’s own visual sensor

3.1.2.2 Music therapy

Music therapy (MT) can transfer skills developed in music-based experiences to other areas of life. Active music composition and musical engagement are valuable for improving attention, memory, and verbal communication in children with ASD. In addition, research has shown that MT effectively reduces anxiety and aggression in people with ASD.(65) Movement sonification can promote the multisensory integration of perception and self-motion.(66) On this basis, a common form of MT uses recognized movement information as a signal to control the music, with the movement itself a critical factor in the performance. Ichinose *et al.*(67) described a novel system that links Kinect and an electronic instrument called Cyber Musical Instrument with Score to provide MT. Magrini *et al.*(68) designed two different versions of the system for controlled and home environments. Image segmentation is used in the version for a controlled environment to obtain a binary human body image, in which a binary raster matrix and tracking algorithm are used to obtain the pose parameters. The version for a home environment uses the Microsoft Kinect Software Development Kit (SDK) library to extract the 3D coordinates of skeletal joints, upon which geometric transformations are performed. Both versions connect the acquired action parameters with the sound parameters. Ragone *et al.*(69) designed an interactive music system that captures the interactive movements of individuals using Kinect v2, then converts them into sound. This system can encourage synchronous movement between autistic children and counselors through an MT environment.

3.1.2.3 Robot-assisted therapy

Robots have become promising tools for aiding rehabilitation and daily skill development in the medical field.(70,71) It has been shown that autistic people can practice life skills more effectively when interacting with robots than with humans.(72) Learning to notice and adequately assess time is a critical first step in improving social skills in children with autism. Ma *et al.*(73) combined MT with robots to estimate the rhythmic cycle of children's movement in a robot-based MT process. To achieve this, they combined recursive neural networks with the fast Fourier transform (FFT), thereby reducing the average offset error and transient delay.

A more common form of robot-assisted therapy is posture imitation, which can foster the development of empathy, one of the most crucial social skills. One of the main features of ASD is a decreased ability to mimic body movements, which is often associated with damage to the mirror nerve cell system.(74) Lidstone *et al.*(75) compared motor imitation scores computed from human observation coding (HOC) methods with those obtained from a fully automated OpenPose 2D computer-assisted movement intervention (CAMI) method and a Kinect 3D CAMI method. It was found that HOC had the lowest discrimination ability and Kinect 3D CAMI had the greatest ability. In addition, some specially designed action feature extraction methods have been designed for this task. Zheng *et al.*(76) developed a rule-based finite state machine to reduce the complexity of computation and the difficulty of generating a training dataset. Guedjou *et al.*(77) proposed a neural network architecture based on a sensorimotor-association paradigm, with visual feature detection based on an attentional vision mechanism committed to sequentially exploring salient points in images. Taheri *et al.*(78) used a state-image-based algorithm and a hidden Markov model combined with a Gaussian mixture model to recognize sequential patterns. Tunçgenç *et al.*(79) designed an algorithm based on metric learning and dynamic time warping that automatically detected and evaluated the critical joints and returned a score by considering the spatial position and timing differences between a child and the model. Ivani *et al.*(80) and Fassina *et al.*(81) used residual neural networks to identify actions, where the former represented an action sequence as an image, retaining the original time dynamic information and spatial structure information, whereas the latter, through the analysis of the subject's kinetic parameters, shaped the beginning and conclusion of each gesture.

To enrich the interaction between robots and users, robots must receive feedback from user actions.(82) Marco *et al.*(83) proposed a technical framework capable of analyzing and integrating multiple visual cues involving face detection, landmark extraction, gaze estimation, head posture estimation, and facial expression recognition, which resulted in accurate head pose estimation by exploiting the information provided by the conditional local neural field. Haibin *et al.*(84) extracted 3D movement trends and geometric properties from upper-body joints to recognize the behavior of children with ASD. Silva *et al.*(85) trained a CNN with different behaviors to classify different behavior patterns using extracted coordinates of the joints of users.

3.1.2.4 Emotional state and engagement during treatment

Engagement is one of the key measures used to assess the impact of therapeutic interventions on children. Poor patient participation may affect training outcomes, especially social skills training for people with autism.(86) On this basis, Dang *et al.*(87) proposed a new classification framework for rehabilitation training activities for autistic children. Motion and electroencephalogram features were integrated into two SVMs to perform frame-based classification for children's motor and psychological assessments. To determine the attention of autistic children attending neurofeedback therapy courses, López-López *et al.*(88) proposed an automated pipeline to obtain head postural features and central position features of children's eyes. As shown in Fig. 1, the CV pipeline took video frames as the input, extracting relevant features through face detection, face recognition, facial vital point detection, head posture evaluation, and eye positioning. Most of the proposed automated methods for subject attention or engagement are similar to this pipeline.

Measuring the child's engagement is crucial to maintaining the interaction of a social robot with the child. Anzalone *et al.*(89) proposed measures to describe a child's behavior in terms of body and head movements, gazing magnitude, gazing direction (left vs front vs right), and kinetic energy. Javed *et al.*(90) and Rudovic *et al.*(91) designed a multimodal child participation model. The former included affective engagement (displayed through eye gaze focus and facial expression) and task engagement (determined by the level of physical activity), whereas the latter used a multimodal audio, video, and autonomic physiology dataset. To further personalize the model for each child, the context layer incorporated demographic variables and an expert-assessed childhood autism rating scale. Although engagement is widely used, the relevant

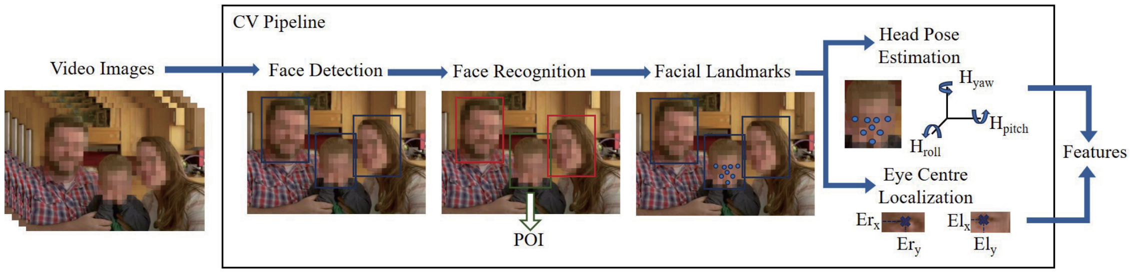

Fig. 1. (Color online) CV pipeline in automatic coding system.[88]

datasets are often small and context-specific. Accordingly, Rakhymbayeva *et al.*[92] used transfer learning to improve the classification accuracy of the Qamqor dataset through the PInSoRo dataset. In addition, how the robot's behavior triggers the child's behavior is essential reference information. Lytridis *et al.*[93] explored this correlation through head pose recognition. They regarded it as a pattern classification problem because they assumed that a child is engaged only if its head is directly oriented toward the robot.

3.2 Relevant datasets

To enable CV technicians to carry out related work, we next summarize the publicly available autism datasets discovered during the investigation.

3.2.1 Social communication and social interaction dataset

Multimodal Dyadic Behavior Dataset (MMDB).[94] This is a collection of multimodal (video, audio, and physiological) recordings of infants' and toddlers' social and communicative behaviors during semistructured play with an adult. This dataset contains the children's social attention, interaction, and nonverbal communication.

3.2.2 Self-stimulatory behavior datasets

(1) Self-Stimulatory Behavior Dataset (SSBD).[95] This unstructured dataset contains home videos of three self-stimulating behaviors in children with autism: head banging, spinning, and hand clapping. The authors time-stamped all three behaviors in the videos. The dataset was extracted from public domain videos posted on video-sharing sites such as YouTube, Vimeo, and Dailymotion.
(2) Expanded Stereotype Behavior Dataset (ESBD).[34] This dataset was collected from public social media channels and consists of videos demonstrating four behaviors. Compared with SSBD, this dataset includes one more behavior (finger movement), and contains almost twice the number of videos. In the two datasets, there are no identical videos.
(3) 3D-Autism Dataset (3D-AD).[96] This dataset is the first 3D dataset available online for research on the 3D recognition of complex and repetitive behaviors of autistic people. This dataset contains these actions (from simple to complex): hands on the face, hands back,

tapping ears, head banging (or rocking back and forth), flicking, hands stimming, hand moving front of the face, toe walking, walking in circles, and playing with a toy from/to different positions repetitively. Each action has been repeated for at least 10 times with non-autistic people.

(4) YouTube ASD.[32] The videos in this dataset were collected from publicly available files on the YouTube video platform. The videos focus on stimming behaviors, which are often fast and atypical, such as clapping, spinning, jumping or swaying back and forth, repetitive play, and fiddling with toys/objects. This database can be accessed via the original post on the YouTube platform and provides the start and end frame numbers for each selected sequence.

3.2.3 Motor deficit datasets

(1) Autism Spectrum Disorder Detection Dataset (ASDD).[55] The dataset comprises 1837 video recordings of children with ASD and TD children, showcasing four different actions: placing, pouring, passing to pour, and passing to place. Each video was shot from a side view using a camera. The video sequence of the moment the hand grabs a bottle is precisely trimmed, with the adjacent sections removed. The dataset was designed to classify pathological and healthy subjects by their differences in performing simple motor behaviors.

(2) Prospective Motor Control in Autism Dataset.[97] This dataset is similar to the previous one. A near-IR camera motion capture system was used to track and record grab kinematics, with six cameras placed 1.5–2 m from the table.

(3) Gait and Full Body Movement Dataset of Autistic Children.[98] The creators of the dataset aimed to diagnose AD on the basis of gait and body movement analysis. Kinect v2 was used to create a 3D dataset, which includes 3D joint positions, a joint trajectories video, a skeleton movement video captured by Kinect v2, and color videos captured by a Samsung Note 9 camera.

(4) 19-Gestures Dataset.[80] This dataset contains gestures from 22 subjects (nine healthy children and 13 adults), three of whom have ASD. The raw dataset was manually segmented and split into training, validation, and test sets through a leave-P-out subject cross-validation to isolate each gesture.

3.2.4 Robot-assisted therapy datasets

(1) Multi-Modal Dataset of Children with Autism (MDCA).[99] The dataset includes (1) video recordings of facial expressions, head and body movements, and gestures of children, (2) the autonomic physiology (heart rate, electrodermal activity, and body temperature) of the children, and (3) audio recordings. The data are from 35 children with different cultural backgrounds.

(2) DE-ENIGMA Dataset.[100] This dataset is a multimodal (e.g., audio, video, and depth) database of recordings of 803 utterances from 14 autistic children aged 4–10 years during Wizard-of-Oz interactions with a humanoid robot. Experts annotated information regarding emotional valence, arousal, audio features, and body gestures.

(3) DREAM Dataset.(101) In this dataset, half of the children interacted with the social robot NAO, while the other half interacted directly with a therapist. Both groups followed the applied behavior analysis protocol. The publicly available version of the dataset comprises body motion, head position and orientation, and eye gaze variables, all specified as 3D data in a joint frame. In addition, metadata containing participants' age, gender, and autism diagnosis variables are included.

3.2.5 Engagement (in MT) dataset

(1) Multimodal Synchrony Dataset (M-MS).(102) For this dataset, multimodal data from a total of 41 sessions (578 min) of MT were collected. The data include electrocardiographic signals, video recordings, behavioral coding, and participants' information. To reflect the different prevalences of ASD according to gender, the study involved 19 male and two female autistic children.

The datasets mentioned above cover various behaviors and application scenarios of ASDs, providing data support for CV researchers lacking medical assistance. Among them, five datasets(94,98,100–102) provide multimodal information, which aids in expanding research content. Three datasets(96,98,101) contain joint 3D coordinates, which help simplify the workflow. Two datasets(32,95) were collected from public platforms, contributing to enhanced system robustness. The MDCA dataset(99) contains data on participants from different cultural backgrounds, helping improve the generalizability of models. However, owing to the ongoing development of research methods, the datasets cannot be generalized to all studies.(103) Furthermore, the subjects exhibit distinct individual characteristics, and research results achieved with these datasets require further verification before their practical application.(104)

4. Discussion

4.1 Research status

We have systematically reviewed the indexed literature from January 2015 to March 2023 on the general use of CV technology based on motion information collected by noncontact devices in autism research. We found that the studies employing noncontact vision sensors have an extensive range of application prospects, and there have been many outstanding works on diagnosis and treatment. Such sensors can also obtain accurate information, such as motion speed, acceleration, and symmetry, through postprocessing.

(1) **Notable research directions**. In work related to the diagnosis of autism, insufficient attention has been paid to motor defects. Although motor deficits are features associated with autism in the medical field, they are correlated with core autism symptoms and broader functions during the entire development of ASDs. Regarding autism treatment and assistance, humanoid social robots are the most promising assistance tool, among which NAO is a commonly used robot, which can be employed not only to cultivate the social communication ability of ASD children but also in postural imitation.

We also found three noteworthy directions of research in our literature review: (1) A multimodal approach.[91] Autism is a complex condition, and a multimodal approach achieves better results than a single-mode approach by combining the knowledge of different modes. Standard multimodal information includes video, voice, physiological signals, and expert scores. Moreover, demographic variables such as patient ethnicity and gender are also helpful. (2) Multiple scenes.[32] Currently, most methods based on CV have specific requirements on the scene, and technologies robust to the background and other interfering factors can be extended to home applications to facilitate the treatment and monitoring of patients. (3) Multiperson tracking.[38,39] A multiperson tracking framework can be used in human intervention therapy or free scenes.

(2) **Feature extraction**. For data collected from contact-free devices, most research has used publicly available tools such as OpenCV, Microsoft Kinect SDK, and an MEA to extract further information. The former two combine CV and deep learning and automatically detect joint coordinates in the human body, namely, they can be used as posture estimation tools. OpenCV includes OpenFace, used to extract facial key points, and OpenPose, used to extract key body points. Microsoft Kinect SDK is employed to extract key points of the body with higher data quality and accuracy than OpenPose.[69] An MEA uses the differential frame technique: it extracts time series motion data using pixel changes in a plane or region. In addition, two interesting findings have been obtained from skeleton-based studies. In research on recognizing atypical behavior, the key points of the head and neck are often removed. People (particularly children with ASDs) tend to look around during treatment regimens, and this movement degrades the performance of the recognition algorithm.[81] However, in the study of atypical movement patterns, the focus is often on the key points of the head. It has been proposed that head movement may potentially provide a new objective biomarker for ASD.[53]

(3) **Experimental data**. Most studies obtained data through custom experiments. The experiments on diagnosis generally referred to an authoritative protocol design, while the experiments on therapeutic work tended to test products and had less explicit description of the reference protocol. Some studies collected data on public platforms, such as YouTube. However, publicly collected data generally face the problem of variable quality. A typical exclusive annotation strategy may reduce the tagging rate when annotations are subjective. Accordingly, Li *et al.*[23] allowed the second and third labels for each instance, namely, they employed the uncertainty-preserved annotation approach. In addition, some work used existing datasets to train models, but these datasets often need to be revised. In this regard, data enhancement,[35] weak supervision,[40] and transfer learning have been attempted.[92]

4.2 Open challenges and future perspectives

(1) **Improve the robustness of suitable CV methods**. The ambiguity of identifying movements originates from the difficulty of customizing body part movements and many other real-world problems, such as camera movements, dynamic backgrounds, and severe weather conditions.[105] Therefore, the high requirement of existing technology in terms of data quality makes data acquisition complex and limits the flexibility of applications.

(2) **Improve the interpretability of the deep learning algorithm**. One of the characteristics of deep learning is black-box reasoning, which makes the detection of two problems difficult: (1) incorrect data are input during model training, leading to incorrect model construction, and (2) the model makes predictions on the basis of training data or prior knowledge, but unseen or problem samples result in incorrect predictions. Markus *et al.*[106] introduced some explainable AI methods and proposed a framework for selecting the most suitable one.

(3) **Diagnose autistic people on the basis of multiple symptoms**. Autism is a complex condition, and ASD is only diagnosed when the characteristic deficits in social communication are accompanied by excessively repetitive behaviors, limited interests, and adherence to the same objects. Autism shares similar features with other neurodevelopmental disorders and often occurs in conjunction with other mental and behavioral disorders that develop in childhood. In addition, the symptoms vary with the progression of the disease and may be masked by compensatory mechanisms.[2]

(4) **Further promote the multimodal fusion approach**. Most studies have focused on RGB data from images or video streams. However, sound and physiological signals contain valuable information for diagnosis. In addition, integrating patient characteristics and demographic informatics will improve the individualized judgment of models.[91] The limitation of multimodal fusion lies in the difficulty of data acquisition and technical design. In addition, multimodal data and sample space reduction need to be balanced because a larger feature space results in higher demands on a system's performance and scalability.[107]

(5) **Improve data sharing**. The lack of publicly available large-scale benchmark datasets is a common problem in healthcare because protecting patients' data is paramount. Most studies involved self-defined experiments and datasets, resulting in no uniform quantitative criteria for comparing results. Therefore, it is necessary to establish standardized experimental conditions and collection methods through the participation of clinical experts, similar to the National Database for Autism Research,[108] which is highly conducive to data sharing.

(6) **Improve the ability to learn from big data**. Research teams lacking medical support should improve their ability to access and use data from public platforms. In data collection, the problem of data imbalance or small samples caused by rare events is common. When there is an imbalance in the class distribution within a dataset, most predictions will align with the majority class. In contrast, features from the minority class will be treated as data noise and consequently be overlooked. As a result, the model will exhibit significant bias.[109]

5. Conclusion

In this review, we explored the research status and prospects for application of motion information obtained by noncontact visual sensors in the intelligent diagnosis and treatment of autistic patients. To ensure the quality and sophistication of the references, we systematically reviewed studies indexed on Web of Science, PubMed, and Engineering Village and published from January 2015 to March 2023. We introduced and analyzed every eligible paper before comprehensively describing the status of research and problems. To facilitate the work of technical personnel, we also summarized the relevant datasets. Our review also has some limitations. Some excellent work may have been excluded due to our research methods. However, to our knowledge, this is the first system in a review of artificial intelligence applications for autism that solely focuses on motion information acquired through noncontact devices.

Acknowledgments

This work was supported by Liaoning Province Higher Education Innovative Talents Program Support Project (Grant No. LR2019058), Liaoning Province Joint Open Fund for Key Scientific and Technological Innovation Bases (Grant No. 2021-KF-12-05), the Central Guidance on Local Science and Technology Development Fund of Liaoning Province (Grant No. 2023JH6/100100066), and Zhejiang Provincial Natural Science Foundation of China (LQ23F030001).

References

1 S. Kalikar, A. Sinha, S. Srivastava, and G. Aggarwal: Proc. 3rd Int. Conf. Communication, Computing and Electronics Systems (Springer, Singapore, 2022) 844. https://doi.org/10.1007/978-981-16-8862-1_66
2 American Psychiatric Association: Diagnostic And Statistical Manual of Mental Disorders, Text Revision (American Psychiatric Association, Washington, 2022) 5th ed., pp. 57–63. https://doi.org/10.1176/appi.books.9780890425787
3 Z. Zhao, H. Tang, C. Alviar, C. T. Kello, X. Zhang, X. Hu, X. Qu, and J. Lu: Autism Res. **15** (2022) 305. https://doi.org/10.1002/aur.2646
4 M. J. McCarty and A. C. Brumback: Semin. Pediatr. Neurol. **38** (2021) 100897. https://doi.org/10.1016/j.spen.2021.100897
5 E. T. Sadek, N. A. Seada, and S. Ghoniemy: Proc. 2020 15th Int. Conf. Computer Engineering and Systems (ICCES, 2020) 1–6. https://doi.org/10.1109/ICCES51560.2020.9334560
6 H. Albert, M. Arnold, and F.-F. Li: Nature **585** (2020) 193. https://doi.org/10.1038/s41586-020-2669-y
7 J. Yu, H. Gao, J. Sun, W. Yang, Y. Jiang, and Z. Ju: IEEE Sens. J. **21** (2021) 11476. https://doi.org/10.1109/JSEN.2020.3017737
8 J. Yu, H. Gao, J. Sun, D. Zhou, and Z. Ju: IEEE Trans. Cognit. Dev. Syst. **14** (2022) 1574–1583. https://doi.org/10.1109/TCDS.2021.3124764
9 R. A. J. de Belen, T. Bednarz, A. Sowmya, and D. D. Favero: Transl. Psychiatry **10** (2020) 225. https://doi.org/10.1038/s41398-020-01015-w
10 E. T. Sadek, N. A. Seada, and S. Ghoniemy: Comput. Sci. **20** (2020) 89. https://doi.org/10.21608/ijicis.2020.46360.1034
11 G. Thomas, A. Dominique, C. Mohamed, C. David, J. Wafa, and A. S. Maria: Cognit. Comput. **14** (2022) 624. https://doi.org/10.1007/s12559-021-09940-8
12 Z. Wang, K. Xu, and H. Liu: Proc. the 13th Int. Conf. Distributed Smart Cameras Conf. (2019) 1–6. https://doi.org/10.1145/3349801.3349826
13 H. Qin, Z. Wang, J. Liu, Q. Xu, H. Li, X. Xu, and H. Liu: Int. J. Intell. Rob. Appl. **13015** (2021) 177. https://doi.org/10.1007/978-3-030-89134-3_17
14 W. Liu, T. Zhou, C. Zhang, X. Zou, and M. Li: Proc. Seventh Int. Conf. Affective Computing and Intelligent Interaction (ACII, 2017) 178–183. https://doi.org/10.1109/ACII.2017.8273597
15 K. Campbell, C. Kimberly, H. Jordan, E. Steven, M. Samuel, S. B. Jana, C. Zhuoqing, Q. Qiang, V. Saritha, A. Elizabeth, T. Mariano, E. Helen, B. Jeffery, S. Guillermo, and D. Geraldine: Autism **23** (2018) 619. https://doi.org/10.1177/1362361318766247
16 S. Perochon, M. D. Martino, R. Aiello, J. Baker, K. Carpenter, Z. Chang, S. Compton, N. Davis, B. Eichner, S. Espinosa, J. Flowers, L. Franz, M. Gagliano, A. Harris, J. Howard, S. H. Kollins, E. M. Perrin, P. Raj, M. Spanos, B. Walter, G. Sapiro, and G. Dawson: J. Child Psychol. Psychiatry **62** (2021) 1120. https://doi.org/10.1111/jcpp.13381
17 J. Hashemi, G. Dawson, K. L. H. Carpenter, K. Campbell, Q. Qiu, S. Espinosa, S. Marsan, J. P. Baker, H. L. Egger, and G. Sapiro: Proc. 2021 IEEE Transactions on Affective Computing (IEEE, 2021) 215. https://doi.org/10.1109/taffc.2018.2868196

18 Z. Wang, J. Liu, K. He, Q. Xu, X. Xu, and H. Liu: 2021 IEEE Trans. Ind. Inf. **17** (2021) 587. https://doi.org/10.1109/TII.2019.2958106
19 C. Song, S. Wang, M. Chen, H. Li, F. Jia, and Y. Zhao: Displays **76** (2023) 102360. https://doi.org/10.1016/j.displa.2022.102360
20 J. Liu, Z. Wang, K. Xu, B. Ji, G. Zhang, Y. Wang, J. Deng, Q. Xu, X. Xu, and H. Liu: IEEE Trans. Cybern. **52** (2022) 3914. https://doi.org/10.1109/TCYB.2020.3017866
21 Y. Shi, W. Ren, W. Jiang, Q. Xu, X. Xu, and H. Liu: Int. J. Intell. Rob. Appl. **13455** (2022) 370. https://doi.org/10.1007/978-3-031-13844-7_36
22 M. Bovery, G. Dawson, J. Hashemi, and G. Sapiro: IEEE Trans. Affective Comput. **12** (2021) 722. https://doi.org/10.1109/TAFFC.2018.2890610
23 J. Li, A. Bhat, and R. Barmaki: Proc. 2021 Int. Conf. on Multimodal Interaction (2021) 397. https://doi.org/10.1145/3462244.3479891
24 A. L. Georgescu, J. C. Koehler, J. Weiske, K. Vogeley, N. Koutsouleris, and C. Falter-Wagner: Front. Rob. AI **6** (2019) 132. https://doi.org/10.3389/frobt.2019.00132
25 Y.-S. Lin, S. S.-F. Gau, C.-C. Lee: IEEE J. Sel. Top. Signal Process. **14** (2020) 299. https://doi.org/10.1109/JSTSP.2020.2970578
26 N. Kojovic, S. Natraj, S. P. Mohanty, T. Maillart, and M. Schaer: Sci. Rep. **11** (2021) 15069. https://doi.org/10.1038/s41598-021-94378-z
27 C. Tang, W. Zheng, Y. Zong, N. Qiu, C. Lu, X. Zhang, X. Ke, and C. Guan: IEEE J. Sel. Top. Signal Process. **28** (2020) 2401–2410. https://doi.org/10.1109/TNSRE.2020.3027756.
28 J. Maha, M. Aicha, M. Djamal, A. Rachid, and Z. Arsalane: Int. J. Biomed. Eng. Technol. **29** (2019) https://doi.org/10.1504/IJBET.2019.097621
29 J. Yu, H. Gao, Y. Chen, D. Zhou, J. Liu, and Z. Ju: IEEE Trans. Hum.-Mach. Syst. **52** (2022) 784. https://doi.org/10.1109/THMS.2022.3144951
30 J. Yu, H. Gao, Y. Chen, D. Zhou, J. Liu, and Z. Ju: IEEE Trans. Cognit. Dev. Syst. **14** (2022) 1654. https://doi.org/10.1109/TCDS.2021.3131253
31 V. Kathan, M. Rui, R. Behnaz, L. Shuangjun, N. Michael, P. Thomas, O. Ronald, and O. Sarah: 2019 IEEE 29th Int. Workshop Machine Learning for Signal Processing (MLSP, 2019) 1st ed., pp. 1–6. https://doi.org/10.1109/MLSP.2019.8918863
32 A. Cook, B. Mandal, D. Berry, and M. Johnson: Proc. 2019 IEEE Int. Conf. Data Science and Advanced Analytics (IEEE, 2019) 504–510. https://doi.org/10.1109/DSAA.2019.00065
33 Y. Tian, X. Min, G. Zhai, and Z. Gao: Proc. 2019 IEEE Int. Conf. Multimedia and Expo (IEEE, 2019) 272–277. https://doi.org/10.1109/ICME.2019.00055
34 F. Negin, B. Ozyer, S. Agahian, S. Kacdioglu, and G. T. Ozyer: Neurocomputing **446** (2021) 145. https://doi.org/10.1016/j.neucom.2021.03.004
35 P. Washington, A. Kline, O. C. Mutlu, E. Leblanc, C. Hou, N. Stockham, K. Paskov, B. Chrisman, and D. Wall: Proc. Extended Abstracts of the 2021 CHI Conf. Human Factors in Computing Systems (2021) 1–7. https://doi.org/10.1145/3411763.3451701
36 L. d. J. Hoekstra, S. v. d. Steen, and R. Cox: Acta Psychol. **211** (2020) 103187. https://doi.org/10.1016/j.actpsy.2020.103187
37 D. Zhang, C. M. Toptan, G. Zhang, S. Zhao, D. Zhou, and H. Liu: Proc. 2021 13th Int. Conf. Advanced Computational Intelligence Conf. (ICACI, 2021) 286–292. https://doi.org/10.1109/ICACI52617.2021.9435864
38 Y. Zhang, Y. Tian, P. Wu, and D. Chen: Sensors **21** (2021) 411. https://doi.org/10.3390/s21020411
39 J. B. S. D. Pandian, S. S. Rajagopalan, D. B. Jayagopi: Proc. 2022 IEEE Int. Conf. Image Processing (IEEE, 2022) 3356–3360. https://doi.org/10.1109/ICIP46576.2022.9897867
40 P. Pandey, P. AP, M. Kohli, and J. Pritchard: Proc. AAAI Conf. Artificial Intelligence (AAAI, 2020) 463–470. https://doi.org/10.1609/aaai.v34i01.5383
41 B. Prabha, M. Priya, N. R. Shanker, and E. Ganesh: Biomed. Signal Process. Control **70** (2021) 103038. https://doi.org/10.1016/j.bspc.2021.103038
42 M. Y. O. Camada, J. J. F. Cerqueira, and A. M. N. Lima: Appl. Soft Comput. **99** (2021) 106877. https://doi.org/10.1016/j.asoc.2020.106877
43 C. J. Zampella, L. A. L. Wang, M. Haley, A. G. Hutchinson, and A. d. Marchena: Curr. Psychiatry Rep. **23** (2021) 64. https://doi.org/10.1007/s11920-021-01280-6
44 A. Caruso, L. Gila, F. Fulceri, T. Salvitti, M. Micai, W. Baccinelli, M. Bulgheroni, and M. L. Scattoni: Brain Sci. **10** (2020) 379. https://doi.org/10.3390/brainsci10060379
45 D. Hirokazu, I. Naoya, F. Akira, S. Zu, Y. Rikuya, S. Kazuyuki, I. Mayuko, S. Koji, and T. Toshio: Sci. Rep. **12** (2022) 18045. https://doi.org/10.1038/s41598-022-21308-y

46 X. Jin, H. Zhu, W. Cao, X. Zou, and J. Chen: Sci. Rep. **13** (2023) 3471. https://doi.org/10.1038/s41598-023-30628-6
47 A. R. Mariano, J. Marín-Morales, M. E. Minissi, G. T. Garcia, L. Abad, and I. A. C. Giglioli: J. Clin. Med. **9** (2020) 1260. https://doi.org/10.3390/jcm9051260
48 A. B. Cho, K. Otte, I. Baskow, F. Ehlen, T. Maslahati, S. Mansow-Model, T. Schmitz-Hübsch, B. Behnia, and S. Roepke: Sci. Rep. **12** (2022) 7670. https://doi.org/10.1038/s41598-022-10760-5
49 N. A. Khan, M. A. Sawand, M. Qadeer, A. Owais, S. Junaid, and P. Shahnawaz: Int. J. Comput. Sci. Netw. Secur. **17** (2017) 256. https://paper.ijcsns.org/07_book/201704/20170435.pdf
50 K. B. Martin, Z. Hammal, G. Ren, J. F. Cohn, J. Cassell, M. Ogihara, J. C. Britton, A. Gutierrez, and D. S. Messinger: Mol. Autism **9** (2018) 14. https://doi.org/10.1186/s13229-018-0198-4
51 G. Dawson, L. Campbell, J. Hashemi, S. J. Lippmann, V. Smith, K. Carpenter, H. Egger, S. Espinosa, S. Vermeer, J. Baker, and G. Sapiro: Sci. Rep. **8** (2018) 17008. https://doi.org/10.1038/s41598-018-35215-8
52 P. R. K. Babu, J. M. D. Martino, Z. Chang, S. Perochon, R. Aiello, K. L. H. Carpenter, S. Compton, N. Davis, L. Franz, S. Espinosa, J. Flowers, G. Dawson, and G. Sapiro: Child Psychol. Psychiatry **64** (2023) 156. https://doi.org/10.1111/jcpp.13681
53 Z. Zhao, Z. Zhu, X. Zhang, H. Tang, J. Xing, X. Hu, J. Lu, Q. Peng, and X. Qu: Autism Res. **14** (2021) 1197. https://doi.org/10.1002/aur.2478
54 J. Yu, Y. Xu, H. Chen, and Z. Ju: IEEE Trans. Neural Networks Learn. Syst. (2022) 1. https://doi.org/10.1109/TNNLS.2022.3216084
55 Z. Andrea, M. Pietro, C. Andrea, A. Caterina, P. Jessica, B. Francesca, V. Edvige, B. Cristina, and M. Vittorio: Proc. 2018 24th Int. Conf. Pattern Recognition (2018) 3421–3426, https://doi.org/10.1109/ICPR.2018.8545095
56 S. Pandya, S. Jain, and J. P. Verma: Proc. 2022 IEEE Bombay Section Signature Conf. (IEEE, 2022) 1–6. https://doi.org/10.1109/IBSSC56953.2022.10037438
57 K. Sun, L. Li, L. Li, N. He and J. Zhu: Proc. ICASSP 2020 - 2020 IEEE Int. Conf. Acoustics, Speech and Signal Processing (IEEE, 2020) 3387–3391. https://doi.org/10.1109/ICASSP40776.2020.9054641
58 G. Shushma, I. J. Jacob: Proc. 2022 Second Int. Conf. Artificial Intelligence and Smart Energy (2022) 1–5. https://doi.org/10.1109/ICAIS53314.2022.9743011
59 J. M. Laura, P. C. Inmaculada, C. R. Pilar, D. O. Isaac, M. Manon, B. G. Enrique, P. S. Alejandro: J. Autism and Dev. Disord. **52** (2022) 169. https://doi.org/10.1007/s10803-021-04934-9
60 S. Piana, C. Malagoli, M. C. Usai and A. Camurri: IEEE Trans. Affective Comput. **12** (2021) 1045–1054. https://doi.org/10.1109/TAFFC.2019.2916023
61 X. Ma and J. Yang: Mobile Inf. Syst. **2021** (2021) 6020208. https://doi.org/10.1155/2021/6020208
62 Q. Wang, X. Wang, and L. Xu: J. Healthcare Eng. **2022** (2022) 4516005. https://doi.org/10.1155/2022/4516005
63 K. P. Ahlers, T. P. Gabrielsen, D. Lewis, A. M. Brady, and A. Litchford: Sch. Psychol. Int. **38** (2017) 586–607. https://doi.org/10.1177/0143034317719942
64 W.-T. Tsai, I. Lee, C.-H. Chen: Univ. Access Inf. Soc. **20** (2021) 375. https://doi.org/10.1007/s10209-020-00724-9
65 B. Applewhite, Z. Cankaya, A. Heiderscheit, and H. Himmerich: Int. J. Environ. Res. Public Health **19** (2022) 5150. https://doi.org/10.3390/ijerph19095150
66 T. R. Stanton and C. Spence: Front. Psychol. **10** (2020) 3001. https://doi.org/10.3389/fpsyg.2019.03001
67 T. Ichinose, N. Takehara, K. Matsumoto, T. Aoki, T. Yoshizato, R. Okuno, S. Watabe, K. Sato, T. Masuko, and K. Akazawa: Int. J. Technol. Inclusive Educ. **3** (2016) 938. https://doi.org/10.20533/IJTIE.2047.0533.2016.0120
68 M. Magrini, A. Carboni, O. Salvetti, and O. Curzio: Proc. ICTs for Improving Patients Rehabilitation Research Techniques (Springer, Cham, 2017) 46. https://doi.org/10.1007/978-3-319-69694-2_5
69 G. Ragone, K. Howland, and E. Brulé: Proc. Interaction Design and Children (2022) 1–12. https://doi.org/10.1145/3501712.3529729
70 J. W. Tan, Q. C. Ding, and Z. Y. Bai: Robot **43** (2021) 9. https://robot.sia.cn/CN/10.13973/j.cnki.robot.200023
71 Q. Miao, C. Y. Sun, M. M. Zhang, and K. Y. Chu: Robot **43** (2021) 539–546, 556. https://doi.org/10.13973/j.cnki.robot.200555
72 A. Kouroupa, K. R. Laws, K. Irvine, S. E. Mengoni, A. Baird, and S. Sharma: PLoS One **17** (2022) e0269800. https://doi.org/10.1371/journal.pone.0269800
73 Y.-H. Ma, J.-Y. Lin, S. Cosentino, and A. Takanishi: Proc. 2021 IEEE Int. Conf. Advanced Robotics and Its Social Impacts (IEEE, 2021) 72–77. https://doi.org/10.1109/ARSO51874.2021.9542841
74 S. Sowden, S. Koehne, C. Catmur, I. Dziobek, and G. Bird: Autism Res. **9** (2016) 292. https://doi.org/10.1002/aur.1511
75 D. E. Lidstone, R. Rochowiak, C. Pacheco, B. Tunçgenç, R. Vidal, and S. H. Mostofsky: Res. Autism Spectrum Disord. **87** (2021) 101840. https://doi.org/10.1016/j.rasd.2021.101840

76 Z. Zheng, E. M. Young, A. R. Swanson, A. S. Weitlauf, Z. E. Warren, and N. Sarkar: IEEE Trans. Neural Syst. Rehabil. Eng. **24** (2016) 682–691. https://doi.org/10.1109/TNSRE.2015.2475724.
77 H. Guedjou, S. Boucenna, J. Xavier, D. Cohen, and M. Chetouani: Proc. 2017 26th IEEE International Symp. Robot and Human Interactive Communication (IEEE, 2017) 256–262. https://doi.org/10.1109/ROMAN.2017.8172311
78 A. Taheri, A. Meghdari, and M. H. Mahoor: Int. J. Soc. Rob. **13** (2021) 1125. https://doi.org/10.1007/s12369-020-00704-2
79 B. Tunçgenç, C. Pacheco, R. Rochowiak, R. Nicholas, S. Rengarajan, E. Zou, B. Messenger, R. Vidal, and S. H. Mostofsky: Biol. Psychiatry: Cognit. Neurosci. Neuroimaging **6** (2021) 321. https://doi.org/10.1016/j.bpsc.2020.09.001
80 A. S. Ivani, A. Giubergia, L. Santos, A. Geminiani, S. Annunziata, A. Caglio, I. Olivieri, and A. Pedrocc: Biomed. Signal Process. Control **74** (2022) 103512. https://doi.org/10.1016/j.bspc.2022.103512
81 G. Fassina, L. Santos, A. Geminiani, A. Caglio, S. Annunziata, I. Olivieri, and A. Pedrocchi: Proc. 2022 Int. Conf. Rehabilitation Robotics (2022) 1–6. https://doi.org/10.1109/ICORR55369.2022.9896536
82 J. Yu, H. Gao, D. Zhou, J. Liu, Q. Gao, and Z. Ju: IEEE Trans. Cybern. **52** (2022) 13738. https://doi.org/10.1109/TCYB.2021.3114031
83 D. C. Marco, L. Marco, C. Pierluigi, F. Francesca, S. Letteria, R. Liliana, P. Giovanni, and D. Cosimo: IEEE Trans. Cognit. Dev. Syst. **10** (2018) 993. https://doi.org/10.1109/TCDS.2017.2783684
84 C. Haibin, F. Yinfeng, J. Zhaojie, C. Cristina, D. Daniel, B. Erik, Z. Tom, T. Serge, B. Tony, V. Bram, V. David, R. Kathleen, and L. Honghai: IEEE Sens. J. **19** (2019) 1508–1518. https://doi.org/10.1109/JSEN.2018.2877662
85 V. Silva, F. Soares, C. P. Leão, J. S. Esteves, and G. Vercelli: Sensors **21** (2021) 4342. https://doi.org/10.3390/s21134342
86 C. Li, Z. Rusák, I. Horváth, A. Kooijman, and L. Ji: IEEE Trans. Neural Syst. Rehabil. Eng. **25** (2017) 726. https://doi.org/10.1109/TNSRE.2016.2591183
87 X. Dang, R. Wei, and G. Li: J. Ambient Intell. Human Comput. **8** (2017) 907. https://doi.org/10.1007/s12652-016-0424-x
88 V. R. López-López, L. Escobedo, and L. Trujillo: Expert Syst. **37** (2020) e12572. https://doi.org/10.1111/exsy.12572
89 S. M. Anzalone, J. Xavier, S. Boucenna, L. Billeci, A. Narzisi, F. Muratori, D. Cohen, and M. Chetouani: Pattern Recognit. Lett. **118** (2019) 42. https://doi.org/10.1016/j.patrec.2018.03.007
90 H. Javed and C. H. Park: Front. Rob. AI **9** (2022) 880691. https://doi.org/10.3389/frobt.2022.880691
91 O. Rudovic, J. Lee, M. Dai, B. Schuller, and R. W. Picard: Sci. Rob. **3** (2018) eaao6760. https://doi.org/10.1126/scirobotics.aao6760
92 N. Rakhymbayeva, Z. Balgabekova, M. Nurmukhamed, K. Burunchina, W. Johal, and A. Sandygulova: Proc. 2022 17th ACM/IEEE Int. Conf. Human-Robot Interaction (IEEE, 2022) 1002–1006. https://doi.org/10.1109/HRI53351.2022.9889577
93 C. Lytridis, V. G. Kaburlasos, C. Bazinas, G. A. Papakostas, G. Sidiropoulos, V.-A. Nikopoulou, V. Holeva, M. Papadopoulou, and A. Evangeliou: Sensors **22** (2022) 621. https://doi.org/10.3390/s22020621
94 J. M. Rehg, G. D. Abowd, A. Rozga, M. Romero, M. A. Clements, S. Sclaroff, I. Essa, O. Y. Ousley, Y. Li, C. Kim, H. Rao, J. C. Kim, L. L. Presti, J. Zhang, D. Lantsman, J. Bidwell, and Z. Ye: Proc. IEEE Conf. Computer Vision and Pattern Recognition (IEEE, 2013) 3414–3421. https://doi.org/10.1109/CVPR.2013.438
95 S. Sundar Rajagopalan, A. Dhall, and R. Goecke: Proc. Proc. IEEE Int. Conf. Computer Vision (IEEE, 2013). https://doi.org/10.1109/ICCVW.2013.103
96 O. Rihawi, D. Merad and J.-L. Damoiseaux: Proc. 2017 14th IEEE Int. Conf. Advanced Video and Signal Based Surveillance (IEEE, 2017) 1–6. https://doi.org/10.1109/AVSS.2017.8078544
97 A. Cavallo, L. Romeo, C. Ansuini, J. Podda, F. Battaglia, E. Veneselli, M. Pontil, and C. Becchio: Sci. Rep. **8** (2018) 13717. https://doi.org/10.1038/s41598-018-31479-2
98 A. Ahmed, H. Israa, and R. Yasen: JPCS. **1818** (2021) 12201. https://doi.org/10.1088/1742-6596/1818/1/012201
99 O. Rudovic, J. Lee, L. Mascarell-Maricic, B. W. Schuller, and R. W. Picard: Front. Rob. AI **4** (2017) 36. https://doi.org/10.3389/frobt.2017.00036
100 B. Alice, A. Shahin, C. Nicholas, A. Alyssa, B. Anton, P. Sergey, F. Michael, G. Maurice, and S. Björn: Proc. Interspeech 2017 (2017) 849–853. https://doi.org/10.21437/Interspeech.2017-730
101 E. Billing, T. Belpaem, H. Cai, H.-L. Cao, A. Ciocan, C. Costescu, D. David, R. Homewood, D. H. Garcia, P. G. Esteban, H. Liu, V. Nair, S. Matu, A. Mazel, M. Selescu, E. Senft, S. Thill, B. Vanderborght, D. Vernon, and T. Ziemke: PLoS One **15** (2020) 1. https://doi.org/10.1371/journal.pone.0236939
102 G. Calabrò, A. Bizzego, S. Cainelli, C. Furlanello, and P. Venuti: Progresses in Artificial Intelligence and Neural Systems, E. Anna, F.-Z. Marcos, M. F. Carlo, and P. Eros., Eds. (Springer Singapore, Singapore, 2021) 1st ed., pp. 543–553. https://doi.org/10.1007/978-981-15-5093-5_46

103 Y. J. Erden, H. Hummerstone, and S. Rainey: J. Eval. Clin. Pract. **27** (2021) 485. https://doi.org/10.1111/jep.13527
104 C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King: BMC Medicine **17** (2019) 195. https://doi.org/10.1186/s12916-019-1426-2
105 J. Imen, B. K. Anouar, A. Ihsen, and M. M. Ali: Forensic Sci. Int.: Digital Invest. **32** (2020) 200901. https://doi.org/10.1016/j.fsidi.2019.200901
106 A. F. Markus, J. A. Kors, and P. R. Rijnbeek: J. Biomed. Inf. **113** (2021) 103655. https://doi.org/10.1016/j.jbi.2020.103655
107 M. Mehak and M. Deepti: Arch. Comput. Methods Eng. **29** (2022) 2811. https://doi.org/10.1007/s11831-021-09682-8
108 N. Payakachat, J. M. Tilford, and W. J. Ungar: PharmacoEconomics **34** (2016) 127. https://doi.org/10.1007/s40273-015-0331-6
109 M. T. Ramakrishna, V. K. Venkatesan, I. Izonin, M. Havryliuk, and C. R. Bhat: Entropy **25** (2023) 245. https://doi.org/10.3390/e25020245

About the Authors

Xuna Wang received her B.S. degree in automation from Harbin University of Science and Technology, China, in 2020. She is working on her M.S. project in intelligent systems at Shenyang Ligong University, China. Her research interests include human motion analysis, healthcare robotics, and human–computer interaction. (xuna_emmm666@163.com)

Hongwei Gao received his Ph.D. degree in the field of pattern recognition and intelligent systems from Shenyang Institute of Automation, Chinese Academy of Sciences in 2007. Since September 2015, he has been a professor at the School of Automation and Electrical Engineering, Shenyang Ligong University. Currently, he is the leader of the academic direction for optical and electrical measuring technology and systems. His research interests include digital image processing and analysis, stereo vision, and intelligent computation. He has published more than 60 technical papers in these areas as first author or coauthor. (ghw1978@sohu.com)

Yutong Zhang received her B.S. degree in measurement and control technology and instrumentation from Shenyang Ligong University, China, in 2022. She is working on her M.S. project in intelligent systems at Shenyang Ligong University, China. Her research interests include facial expression analysis, healthcare robotics, and human–computer interaction. (yutongzhang0912@163.com)

Yueqiu Jiang received her Ph.D. degree in computer application technology from Northeastern University in 2004. Since 2010, she has been a full professor at Shenyang Ligong University. Currently, she is the leader of the subject direction for signal and information processing. Her research interests include network management and image processing.
(yueqiujiang@sylu.edu.cn)

Jiahui Yu received his Ph.D. degree in intelligent robotics from the University of Portsmouth, U.K., in 2021. He received his B.S. and M.S. degrees in intelligent systems from Shenyang Ligong University, Shenyang, China, in 2017 and 2019, respectively. He also worked as a research associate at The Chinese University of Hong Kong, Shenzhen, and Shenzhen Institute of Artificial Intelligence and Robotics for Society in 2021. Currently, he is working in the Department of Biomedical Engineering, Zhejiang University, and the Innovation Center for Smart Medical Technologies & Devices, Binjiang Institute of Zhejiang University. His research interests include healthcare robotics, human motion analysis, image processing, and human–robot/computer interaction and collaboration. (jiahui.yu@zju.edu.cn)