IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 1, 2018, accepted August 10, 2018, date of publication August 30, 2018, date of current version September 21, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2867880

Cognition-Based Context-Aware Cloud Computing for Intelligent Robotic Systems in Mobile Education

JIANBO ZHENG[1,2], QIESHI ZHANG[1,2], (Member, IEEE), SHIHAO XU[1,3], HONG PENG[3], AND QIN WU[1,4]

[1]Guangdong Provincial Key Laboratory of Robotics and Intelligent System, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[2]Department of Mechanical and Automation Engineering, the Chinese University of Hong Kong, Hong Kong
[3]School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China
[4]School of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China

Corresponding author: Qieshi Zhang (qs.zhang@siat.ac.cn)

This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2014CB744600, in part by the National Nature Science Foundation of China under Grants 61403365, 61402458, 61632014, 61210010, and 61772508, in part by the Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, under Grant 2014DP173025, in part by the Guangdong Technology Project under Grants 2016B010108010, 2016B010125003, and 2017B010110007, in part by the Natural Science Basic Research Plan in Shaanxi Province of China under Grants 2017JM6101, 2017JQ6077, in part by the Shenzhen-Hongkong Innovative Project under Grant SGLH20161212140718841, in part by the Shenzhen Engineering Laboratory for 3-D Content Generating Technologies under Grant [2017]476, in part by the Shenzhen Technology Project under Grants JCYJ20170413152535587, JSGG20160331185256983, and JSGG20160229115709109, in part by the Program of International S&T Cooperation of MOST under Grant 2013DFA11140, and in part by the Fundamental Research Funds for the Central Universities under Grant GK201703060.

ABSTRACT At present, artificial intelligence (AI) has made considerable progress in recognition of speech, face, and emotion. Potential application to robots could bring significant improvement on intelligent robotic systems. However, limited resource on robots cannot satisfy the large-scale computation and storage that the AI recognition requires. Cloud provides an efficient way for robots, where they off-load the computation too. Therefore, we present a cognition-based context-aware cloud computing framework, which is designed to help robot's sense environments including user's emotions. Based on the recognized context information, robots could optimize their responses and improve the user's experience on interaction. The framework contains a customizable context monitoring system on the mobile end to collect and process the data from the robot's sensors. Besides, it integrates various AI recognition services in the cloud to extract the context facts by analyzing and understanding the data. Once the context data is extracted, the results are pushed back to mobile end for making a better decision in the next interactions. In this paper, we demonstrate and evaluate the framework by a real case, an educational mobile app for English learning. The results show that the proposed framework could significantly improve the interaction and intelligence of mobile robots.

INDEX TERMS Emotion recognition, context-aware, intelligent robotic system, cloud computing, mobile education.

I. INTRODUCTION

With the rapid development of robotics technology, more and more robots take part in people's daily life and assist people on learning and entertaining. One of the typical examples is the commonly used educational robots for kids. Kids can interact with these robots by pressing the buttons with different functions in one way; in another way, the kids' commands can be comprehended by the robots through recognition of voice. For example, kids can ask them to tell stories, sing songs and so on. However, currently most of educational robots for kids lack intelligence since their interactions are based on the preset programs for the same instruction, their responds usually are the same, which causes these robots are not smart enough and become boring soon [1].

Recently, the face and speech recognition has been a leap forward in the development of mobile technology [2], [3].

Meanwhile, the emerging deep learning technology also reinforces the artificial intelligence (AI), which makes it possible for practical context-aware cloud computing that robots could perceive the environment and user's emotions. The context awareness information will help robots understand the human's instructions better and feedback smarter responses.

As known, although it is more reliable and instant to process tasks locally, the mobile device always has its limitation on computing capability. Not all the tasks are suitable for processing locally especially for the heavy computation tasks. Meanwhile, cloud computing is known on providing services with unlimited computing resources and capacities. More and more large-scale AI processing and machine learning have been implemented in the cloud as services [4], such as Google's speech recognition service and face recognition service from Microsoft Azure. Obviously, we can combine mobile computing with cloud computing to achieve more appropriate resources allocation.

Currently, the context-aware computing on mobile devices mostly focuses on the resources off-loading based on computing capacity, network status and energy consumption. In this paper, we propose a cognition-based context-aware cloud computing (C4) framework to help robots understand its runtime context changes and adjust its responses strategy in the interaction based on the context information. Compared with other researches, it aims to continuously optimizing the robots' interaction based on understanding the application scenario, e.g., the age, sex and emotion of the current user, the environment (e.g., in-door or out-door, current weather information) and so on. Based on understanding context information including the recommendations, the proposed framework will help robotic systems improve the interaction, especially on the mobile education area.

In section II, the related researches and technologies will be summarized. Section III shows the overview of the framework, from the context-aware monitoring system on the mobile end to the multi-modal fusion based recognition services in the cloud. Implementation strategies are explained in section IV, which include the design thinking and principles in system's robustness, scalability and user privacy protection. In section V, we demonstrate a real case of C4 framework, conduct an experiment and evaluate the results. Finally, the conclusion and future work are given and discussed in the last section.

II. BACKGROUND AND RELATED WORKS

This section introduces the background and foundation of our work. The existing tools and methodologies we applied in this paper, e.g., voice and facial expression recognition. Related works are summarized to discuss the advantages and disadvantages and show the differences and innovation of our work.

A. AUTOMATIC SPEECH RECOGNITION

At present, the automatic speech recognition (ASR) technology is relatively mature and widely used in voice dialing, voice navigation, indoor device control, voice document retrieval, and simple dictation data entry. Speech recognition has once again achieved a very big breakthrough. Many organizations such as IBM and Microsoft have launched their own Deep CNN models to improve the accuracy of voice recognition. The introduction of the Deep Residual/Highway Network enabled us to deepen the training of neural networks. In addition to the basic semantic recognition requirements, semantic understanding is also very important for a better interactive experience. Semantic comprehension is divided into text semantics and phonetic semantics. It is mainly to convert natural language content into text data with a certain structure, so that the application can grab the key data, understand the intention of the user, and perform the next process. Voice semantics is the first to convert audio data based on speech recognition to natural language text, and the server automatically understands the text semantics. Such as ''What kind of clothes do I wear today?'' This sentence, when the hypothetical content is limited, can directly resolve the user's intention in the application - query the weather in the local area. However, it is difficult for the application to understand the user's intention in each sentence from many texts. When the coverage becomes wider, the amount of time spent on the string-matching calculation become larger on PCs and mobile devices with lower hardware computing speed.

One type of semantic understanding is based on the framework semantics, which is mainly to formally represent the meaning of the text. Each type of forms becomes a framework [5], which generally contains the framework type, framework context, and corresponding parameters. Therefore, each framework accurately represents an event and has different language expressions. Yang and Mitchell [6] proposed a method for generating frame semantics from unstructured texts. They mainly completed the joint model of frame type recognition based on multi-layer neural networks and role tag prediction and predicate-argument span based on long short-term memory (LSTM). Ringgaard *et al.* [7] used Tensorflow and Dragnn to construct a generic transformation-based framework semantic parser that uses BiLSTM for encoding and decodes using recursive units. Compared to semantic role annotation, semantic pre-storage graphs can provide a deeper semantic representation that constructs a factual or logical relationship between real words. Wang *et al.* [8] proposed a Stack LSTM-based transition analyzer to generate dependency graphs and proposed two neural network models: BiLSTM descriptive transfer buffers and subtrees constructed using tree LSTM descriptions.

B. FACIAL EXPRESSION RECOGNITION

In Human-Computer Interaction (HCI) application, gestures, body movement, and facial expressions can convey the feeling and the feedback to the user. The mainstream facial expression recognition method is still based on static pictures. It develops from the methods based on manual extraction

of features, and surface learning, to the present deep learning. Thus, three stages of methods based on deep learning are developed: 1. Preprocessing, 2. Deep feature learning, 3. Facial Expression Classification. Throughout the whole research of facial expression recognition, its development between face recognition promotes each other. This would imply that a better method for face recognition is also applicable to facial expression recognition.

Data sets in this area tend to transfer from the small sample size under the unified control of laboratory to the large and diversified database in real life. And the direction of the algorithm changes greatly, traditional manual design features and even surface learning features can no longer adapt well to various interference factors irrelevant to facial expressions in the real world, such as lighting transformation, different head posture and facial blocking. Therefore, more and more researches began to apply the deep learning to facial expression recognition. There began to be competition of facial expression recognition since 2013, such as FER2013 and EmotiW2016.

For achieving the facial expression recognition, deep learning technologies are utilized in several method. For example, the image-based network adopted Convolutional 3D such as the Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and adapt classifier was used in the network structure for better accuracy. Also, some emotion classifier using deep features which combined with multi deep neural networks in parallel. In general, the latest facial expression recognition methods tend to fuse various deep neural networks.

C. RELATED WORKS

With so many new technologies coming up in the near decades, one of the hottest areas cloud robotics have attracted massive attention of people, since it can serve people in many aspects such as industries, houses, medical and education fields. The robotics in education is a new and promising area especially when the children's future is concerned [9], [10]. Our goal is to construct an intelligent robotic system over clouds with context-awareness in mobile education.

To build the cloud robotic system, two complementary levels are required: machine-to-machine (M2M) connection and machine-to-cloud (M2C) infrastructure [11], [12]. On M2M level, robots connect to each other through wireless links which can share a much larger virtual computing pool compared to a single robot and exchange information within the group. On M2C level, the cloud acts as a shared pool implementing computations, storage and elastic resource allocations for robots' real-time demand. Zhihui Du and his team designed a novel Robot Cloud stack and adopted service-oriented architecture (SOA) which took advantage of the computing of robotics – robot as a service (RaaS) in Cloud Computing and the target business model to make the entire system more flexible, extensible and reusable [13]. Hu *et al.* [14] built a framework called Cloudroid that could support direct deployment of existing robotic software packages to the cloud. That framework functioned as a platform-as-a-service (PaaS) cloud infrastructure carrying transparent service wrapping, which adopts the robot operating system (ROS) package model, and cooperated QoS awareness for effectively maintaining certain QoS properties.

As our serving system aims at the education, interactions between robots and humans become the important part involving such capabilities as activity recognition, semantic reasoning [15]–[17] and facial/emotion recognition so called context-awareness, with which it can not only achieve higher efficiency in responding but also save much more energy. In mobile edge computing (MEC), Sneps-Sneppe and Namiot [18]proposed a cloud-based context-aware computing including content generation, content optimization (ambient mobile intelligence), and smart transportation to target proximity related applications and services. In education, intelligent humanoid robot, a product based on Google translator and basic natural language processing, is doing well on natural interaction for entertainment as well as basic mathematics teaching [19]. While, in industry Wan *et al.* [20] have created an advanced material handling cloud system based on context-aware services and effective load balancing . In emotion recognition, an emotion-aware cognitive system is proposed by Hu *et al.* [21], which provide multidimensional emotional data collection and processing approaches for mobile applications working in mCRAHNs. There are also other kinds of awareness such as the one based on context for mobile crowdsensing in social network architecture [22], the one based on quality embedded in cooperative service access system for social internet of vehicles [23] and the one on situation for delivering music to drivers to ease mood as a crowd-cloud codesign approach [24].

Another concern is computation offloading. It is a way to decrease energy consumption and increase responsiveness for applications by having them to divert computations from locals to remote servers. To investigate the trade-off between energy consumption and latency, an energy-aware offloading scheme was presented by J. Zhang and his team, which jointly optimize communication and computation resource allocation under the limited energy and sensitive latency. To find out the optimal solution, they use an iterative search algorithm combining interior penalty function with D.C. (the difference of two convex functions/sets) programming (IPDC) [25]. To optimize the offloading issues, Liu *et al.* [26] presented a framework for context-aware computation offloading that could help reduce execution time by 6%-96% and power consumption by 60%-96% for computation-intensive applications . Besides, with a mobility model and a fault tolerance mechanism, a novel idea was proposed to overcome problems caused by user movements in the mobile cloud computing context for optimizing offloading decisions [27]. In health-care service area, there are some monitoring applications with context-awareness to achieve a balanced trade-off among energy efficiency, diagnostic accuracy, and processing latency as well [28].

III. SYSTEM DESIGN

In this section, we introduce the architecture of the framework. In general, the purpose of the framework is to monitor the robot's sensor data and understand its runtime context changes. Based on the updated context information, the robot can continuously adjust its responses strategy in the interaction.

As shown in Fig. 1, the framework composes of two main components, context awareness monitoring system and multi-modal based interaction states understanding system. The first component monitors the changes of context and keeps the processed information locally for supporting the robotic interaction. The second component fuses multi-modal emotion recognition results, such as expressions, speech, and gestures aim to obtain interactive more accurate, efficient, and natural.

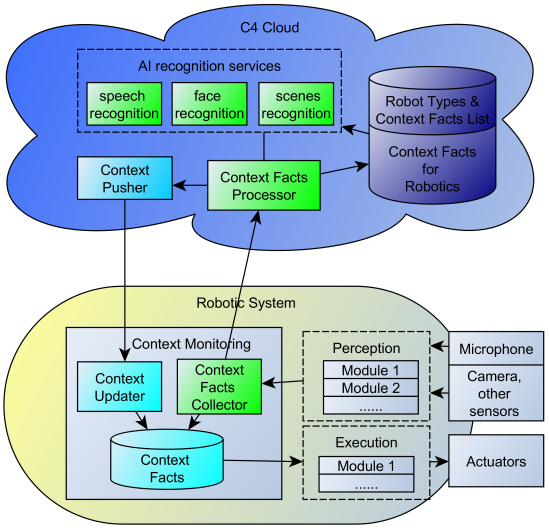


FIGURE 1. Overview of the framework main components.

A. CONTEXT AWARE MONITORING SYSTEM

The monitoring system is designed to continuously collect and analyze the customized facts of context profile locally or in the cloud, to make the robotic system understand its working environment better, including its current user.

In the real scenarios, different scene changes (including background, environment, content of speech, expression, etc.) affect users' interested topics for the interaction. For example, in a process of learning, if the user shows frustration, the system should be able to automatically adjust the difficulty level of the lessons (learning content). After the emotion of user is recognized and understood, more interesting contents will be provided instead of scheduled contents. Another example, for music recommendation, if it happens in the evening before sleep, the recommendation of soft music should be the main theme.

To achieve these personalized recommendations, the robot's reference to the environment include: user's gender, age; current emotion; interested topics; indoor or outdoor; weather conditions and other factors.

1) CUSTOMIZABLE CONTEXT PROFILE

The monitoring system maintains a customizable context profile which consists of the context facts that the robotic system should be monitoring. It keeps monitoring them locally, processing them on the cloud and synchronizing the analysis results back to robots.

The profile includes four categories of context facts, user information, surrounding environment facts, robot itself facts, and other customized facts by the developer.

- User information includes name, gender, age group, emotion and interested topics. The gender, age group, and emotion are getting from the facial recognition. The interested topics usually keep the recommended contents or tasks from cloud services.
- Surrounding environment facts include location, weather, outdoor or indoor for scene recognition.
- Robot itself facts include computation capacity (CPU, RAM), storage capacity, battery info, network status.

The facts in the context profile are customizable. The framework allows to add new facts for different types of robots, as well as the process of collection and analysis.

The framework provides web service interface for the context facts registration, where the robotic developer could define what facts to be monitored and how to process them. By default, the context monitoring system contains a preset of rules for common context facts, and it also allows the developers to add customized rules for various context facts processes.

2) CONTEXT FACTS PROCESSING

Context facts processing is to extract facts of context profile from mobile sensors' raw data. This process includes two stages, one is about collecting the sensors' raw data from mobile end, and the other is about to analyze and understand the collected raw data and get the context awareness data in context profile. The whole process in C4 framework takes place on the mobile end and the cloud end.

- On mobile end

Context facts collector is the main component of C4 framework working on the mobile end to collect the raw data of sensors and handle them based on the rules of context profile.

As shown in Fig. 2, the general process of sensor raw data monitoring is that according to a set interval, the monitor thread of context fact collector periodically detects the raw data of the sensor that listed in the context profile. According to rules in the context profile, when the raw data satisfies the condition of the rule, the preset action will be fired to process it.

Usually, the conditions in the rule are simple. For example, the difference between the current value and the previous is greater than some threshold, or some specific signal is detected, such as the voice activity detection in the sound or human face detection in the camera images.

From the point view of context facts, there are two types of actions to handle the sensor data. One type of action is to

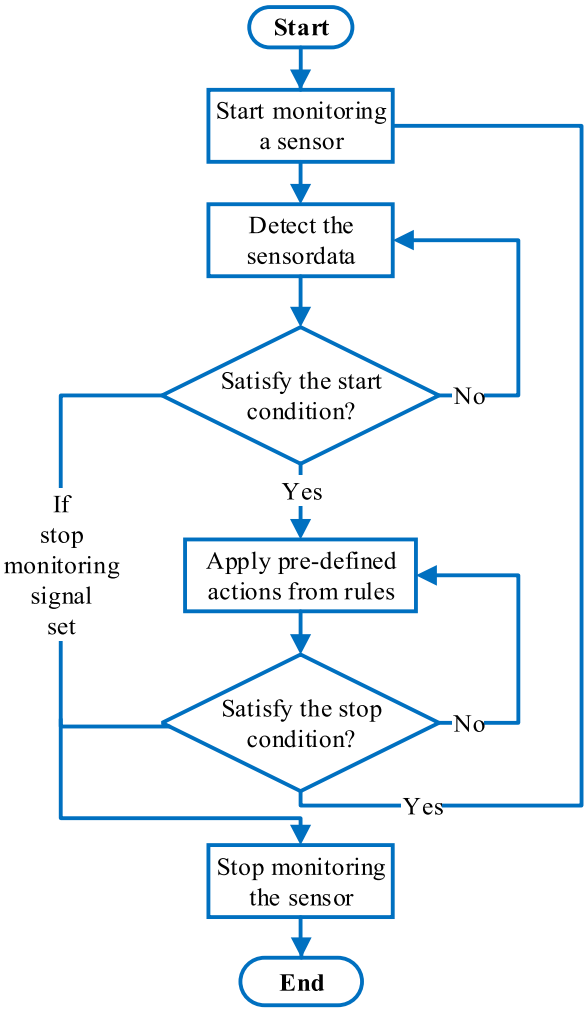


FIGURE 2. The general process of a sensor raw data monitoring.

detect the raw data periodically and then update the context facts. For these types of data, only the recent data need to be stored. For example, the status of mobile, including battery, storage, and network status.

The other type of action is to detect the raw data continuously but only when some condition occurred, the context fact needs to be updated by processing the raw data, such as sound data from microphone or image data from camera. For example, only when user's voice is detected in the sound data, the clip of that voice will be processed and uploaded; only when the human face is detected in the camera, the handler starts to acquire and upload the images from the camera periodically until the human face couldn't be detected for a preset time long. Note that the monitored background sound data is only used for the emotion recognition.

Context updater is responsible for receiving the analysis results from the cloud side and send the notification to the mobile system. All the update actions of local context facts are only executed by the component context updater. When it executes update actions, the corresponding events will be triggered, like a remaining power changed event.

- On cloud end

Context facts processor is the main component of C4 framework working on the cloud end, which is to extract the context facts from the data of mobile end. It provides a set of web services, thought which, it receives various sensor data and processes it into context fact values.

Multiple services including our own and the third-party services are integrated in the context facts processor. There are two ways to integrate these services. For our own recognition services running on our local GPU servers, they are lacking published end points. The context facts processor communicates with them through SSH tunnel, such as the services of face and speech recognition. While for those third-party services on the internet, they are integrated via HTTP request.

These integrated services make the analysis capability of the processor easy to extend. For example, when the GPS data is received by the processor, it will convert the GPS data into the real city first and then fetch the weather information from the weather forecast service providers.

Context pusher is the component that pushes the analysis results so called updated context facts back to the context updater of mobile end. The heart beat is kept between context pusher and updater. It is also monitoring a queue which receives the updated context facts from the processor. Once there are updated context facts came out, the component will push them to the mobile end as soon as possible.

B. MULTI-MODAL FUSION BASED EMOTION RECOGNITION

Emotion recognition can be used to analyze and describe the state of learners, and to adjust the corresponding teaching strategies. However, the identification of single modality is often not accurate enough because of the complexity of the environment and people's intentional or unintentional disguise of their emotions in different states. Therefore, this paper proposes a multi-modality emotion recognition algorithm as Fig. 3 shown, to overcome the inadequacy of existing single-modality algorithms in some situations to accurately reflect the user's emotions, to achieve interactive learning based on emotion recognition, to overcome the lack of emotions of existing online education and other teaching methods that lack of communication and interaction.

Cognitive psychologists have found through many experimental studies that the emotional states which affect human learning cognitive processes include boredom, confusion, happiness, frustration, concentration, and surprise. These emotional states are closely related to people's process of learning cognition. Therefore, these emotional states are called cognitive emotional states. Most of the current emotion recognition researches are based on the analysis of these six basic emotions, while the research on cognitive emotional states is less. Although the speech-based approach is more intuitive, in people-to-people communication, the information passed through the language is often controlled

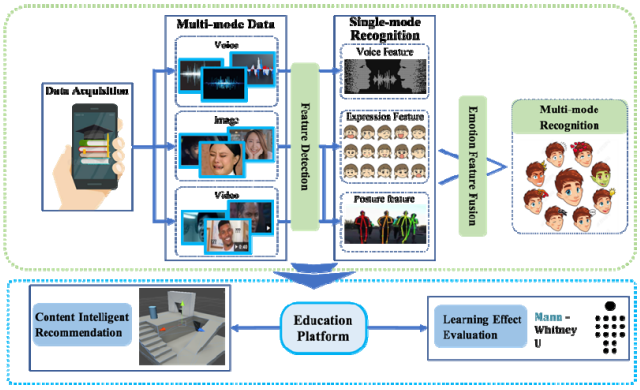


FIGURE 3. Multi-modal emotion recognition based interactive teaching framework.

by rational awareness, and fraud may occur, and the true inner heart cannot be expressed. The non-linguistic behavior is even more expressive in communication and it is close to people's true feelings. Many studies shown that the non-verbal communication can express the true feelings of an individual's heart, as well as many emotions and intentions that are difficult to express in words. Therefore, through the non-linguistic identification analysis, it is possible to obtain more realistic psychological and emotional states in most case. The research of emotion recognition based on expression and limb can reflect the state of emotion more realistically. Combining the characteristics of different modalities often has additional help for emotion recognition. Therefore, with the development of computer and deep learning, scholars at home and abroad began to apply new theories and methods of various disciplines to carry out research on emotion recognition based on multimodal fusion.

In the existing emotion recognition, the single-mode human-machine interaction such as facial expression recognition, speech interaction, and gesture recognition has achieved good results. However, in a virtual learning environment, information such as a person's expression, speech, or gesture alone is used. It is difficult to accurately convey people's true feelings. The multi-modal human-computer interaction technology that integrates speech, expression, and gesture has very important significance for the construction of virtual learning environment. Therefore, this section intends to fuse multi-modal information, such as expressions, speech, and gestures, and break through the deficiencies based on single modality to obtain more interactive, accurate, efficient, and natural. This study includes the following steps:

- Single modal emotion feature extraction
  This step separately extracts speech, expression, and pose information: For speech signals, extracts rhythm features, sound quality features, and MFCC emotion features; for expression signals, RGB-D images are processed. Based on the deep confidence network and the deep residual network, a detailed face model is established and fed back by scanning the facial action units, skin texture changes, and the structure of deformable points; for gesture signals, single frame human body segmentation based on spectral clustering is used. Algorithms, depth-view learning segmentation algorithm for human body parts, motion recognition methods for high-dimensional decision trees, etc. extract various parameters describing pose emotion features to compose pose emotion features.
- Emotional feature fusion
  The eigenvalues extracted from the models are normalized, and then neural network models are merged to concatenate the hidden layers of different modalities. According to the different importance of different modalities for emotional recognition, the multi-modal attention selection mechanism is dynamic. Adjust the weights of different modal fusions, and build a hierarchical structure based on deep neural networks to achieve cross-modal associative learning.
- Multimodal emotion recognition
  Based on the fusion of emotional features, a multi-modal deep learning network is constructed by using attention selection mechanism, and a predictive model is obtained by training the emotional characteristics of the training samples, and cross-modality related learning is implemented through a hierarchical structure to achieve emotional categories. determination.

IV. IMPLEMENTION STRATEGIES

The C4 framework aims to make mobile robots smarter on human-robot interaction by supporting them with AI recognized context-awareness data.

For the prototype of the framework, we chose Java EE application platform with functional REST web services as the server side in the cloud. On the mobile end, from the quantity of devices, android devices probably are the most widely available intelligent robots. Therefore, we chose the android platform as the implementation platform of mobile end. The mobile portion of C4 prototype is implemented as a jar package library on the android platform, which will be easier to distribute and be imported in other mobile apps.

A. DATA TRANSMISSION

For the raw data that is monitored, not all the data will be uploaded. For each type of sensor, the framework allows developer to define the data collection rules, which will define the conditions of start and stop collecting.

On one hand, only when the data meets the collecting condition, it will be uploaded to the cloud end. The condition can be some event happened or the difference is over the preset threshold. For example, for camera image data, only the human face is detected in the camera, the camera image data will start uploading to the cloud periodically.

On the other hand, once the uploading process triggered, it starts checking the stop uploading conditions based on the preset rules. For example, after the human face disappears in the camera for half minute, the context fact monitor will stop uploading the images. This transmission strategy helps saving energy and optimizes the battery life.

B. USER PRIVACY PROTECTION

User privacy protection is always a hot topic and cared by the end users in cloud computing, since the private information is stored in somewhere that they couldn't fully be controlled.

In the implementation of C4 framework, it keeps a user profile to maintain user's personal data, such as user's basic information, face identifier, voice identifier, and interested topics. And all the user's sensitive data will be encrypted in data storage and transmission. The mobile device ID and face recognition result together will be used to identify the user. If the current user doesn't match with any users in the database, the framework will create a new user profile with proper roles for the rest of session.

Except the sensitive user profile, the framework also only saves the analysis results of the context fact. After the platform extracted the context facts from the raw data, it will delete the raw data immediately and won't keep the raw data in the cloud, especially for the images and voice data. By working in this way, it is not only reduced the storage usage in the cloud, but also protected the user's privacy better.

C. CONTINUOUS RECOMMENDATIONS

During the context monitoring, the C4 framework keeps pushing the recommended tasks and topics to the mobile end via context fact ''interested topics''. The mobile end will apply the recommendations in the following interactions with current user.

In C4 framework, there are two types of interested topics, long-term interested topics and short-term interest topics.

The former is extracted from user's habits in daily usage and keywords' frequencies in the voice recognition. These topics are stored in user's profile as interested topics. It is more stable than the latter one, so called long-term interested topics.

The latter is extracted from the combination of user's long-term interested topics, user's current emotion, and the current mobile running task. It is stored in the context profile as ''interested topics''. Usually, the short-term interests consist of the tasks and contents that user may interested. The mobile end can execute the related tasks in the following interactions. The short-term interested topics in context profile will be continuously updated based on the changes of user's emotion and running tasks. For example, if the user is showing bored or tired emotions in the learning, the short-term interested topics will contain the contents that make user excited, such as play excited music or video that user interested in.

In the implementation of C4 framework, we demonstrate the feasibility of its recommendation mechanism and keep the recommendation service interface there. The algorithm of recommendation is more related with the specific reality area. According to the robots' functionality, for different mobile robots, the recommendation strategy should be different.

V. CASE STUDY AND EVALUATION

In this section, a case study of C4 framework in the real world will be demonstrated. And based on results of an experiment, the evaluation will also be given by the comparison before and after applying this framework to an educational app.

A. CASE STUDY

With C4 framework, we developed an android app named Smart Study, which is aiming to help children learn English words, as shown in Fig.4. Most of English daily words for elementary school have been built-in the app.

FIGURE 4. The user interfaces of the app Smart Study.

Before exposing it to kids to learn words, the app allows the teacher or parents to pick up words as a list for kids to learn and the words will be displayed with corresponding pictures as shown in Fig.4 (a). In Fig.4 (b), only the correct picture is selected, the next words will show up. Kids can practice the new learnt here.

To apply the C4 framework to the development of android app is simple. First, developers need to obtain an API key for the app from the cloud site, and customize the context profile for the monitor in the app. With the profile, developers can set the context facts which need to be collected and how to collect them, e.g., the protocols of sensors' data.

After all set on the cloud, in the app project, the library file of the framework needs to be imported and configured with the correct information such as API key and service IP address. In the development, after the app starts, a context monitor will be instantiated and starts monitoring according to the context profile. When a context fact changed, the corresponding event will be triggered. The developer can retrieve the latest context fact from the database and execute corresponding functions to handle the context change event. To make the monitor of C4 framework working, the related

privileges should be assigned to the app, such as accessing to the sensors of camera, microphone and GPS location.

By tracking the updates of user's emotion, interest topics and device usage scene via C4 framework, smart study app can adjust its teaching content and its difficulty level. For example, during the learning process, if the emotion of the user become tried or not focused on the app, the following teaching content will contain some music or animation clips to attract his/her attention back. And these music and animation clips are chosen based on the interested topics in the context profile.

B. EVALUATION

In this section, an experiment with 20 elementary students is carried out. The experiment has two rounds. In the first round, the app is without C4 framework, and in the second round, the app is with C4 framework.

In the experiment, the students are asked to use the app to learn 10 English words for 20 minutes every work day in a week. A quiz would be given after that. During the experiments, the attention time of students learning on the app and scores of the quiz are recorded.

The attention time is calculated based on students' faces and their operations. If the face is shown in front of the device and the app also receives interaction with in 30 seconds, this half minute will be counted in the attention time. Otherwise, the student is considered as distracted.

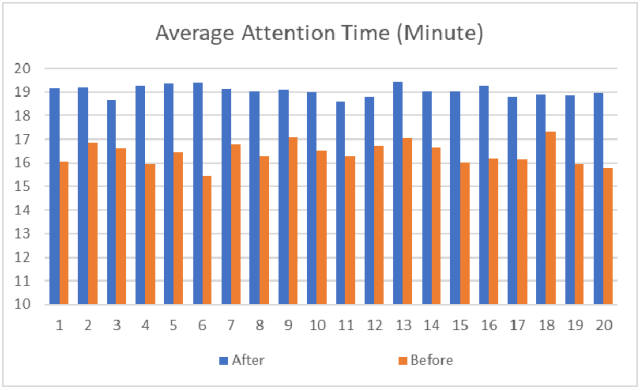


FIGURE 5. The average attention time of students during learning English words before and after C4 framework applied in the app.

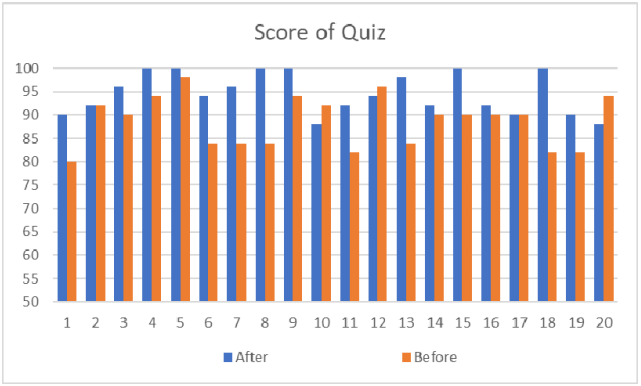


FIGURE 6. The scores of quizzes of students before and after C4 framework applied in the app.

As the results shown in Figs. 5 and 6, we can see that, in overall, after applied C4 framework to the smart study app, the children have more attention time on the English study and the scores of quizzes are also improved significantly. The app with C4 framework has a dramatic improvement on the effect of learning process.

So, from the results, we can see the C4 framework could significantly improve the interaction and intelligence of mobile robots, especially on the mobile education area.

VI. CONCLUSION AND FUTURE WORK

In this paper, we present a framework of cognition-based context-aware cloud computing for intelligent robotics systems in mobile education. First, the continuous monitoring system for customizable context is present, which is the bridge connecting the AI recognition services and the mobile robots. Second, a multi-modal fusion based emotion recognition is proposed to fuse multimodal emotion information, such as facial expression and speech and so on.

Then we apply the framework to support an English educational app and the evaluation results show that the framework can help significantly improve the app's interaction and intelligence.

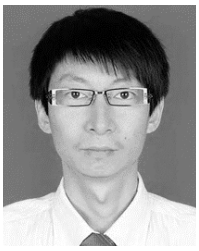
In future, on one hand, we will continue working on the stability and scalability of C4 framework, as well as the energy saving strategies on the mobile end. On the other hand, we will try to apply more machine learning technologies on the cloud end to continuously improve the emotion recognition and the data analysis on the effects of mobile robots' interaction.

REFERENCES

[1] M. E. Karim, S. Lemaignan, and F. Mondada, "A review: Can robots reshape K-12 STEM education?" in *Proc. IEEE Int. Workshop Adv. Robot. Social Impacts (ARSO)*, Jul. 2015, pp. 1–8.
[2] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2016, pp. 1–10.
[3] D. Amodei *et al.*, "Deep speech 2: End-to-end speech recognition in English and mandarin," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2016, pp. 173–182.
[4] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. U. Khan, "The rise of 'big data' on cloud computing: Review and open research issues," *Inf. Syst.*, vol. 47, pp. 98–115, Jan. 2015.
[5] C. F. Baker, C. J. Fillmore, and J. B. Lowe, "The Berkeley framenet project," in *Proc. 17th Int. Conf. Comput. Linguistics (COLING)*, vol. 1, 1998, pp. 86–90.
[6] B. Yang and T. Mitchell, "A joint sequential and relational model for frame-semantic parsing," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2017, pp. 1247–1256.
[7] M. Ringgaard, R. Gupta, and F. C. N. Pereira, "SLING: A framework for frame semantic parsing," *CoRR*, 2017.
[8] Y. Wang, W. Che, J. Guo, and T. Liu, "A neural transition-based approach for semantic dependency graph parsing," in *Proc. 32nd AAAI Conf. Artif. Intell. (AAAI)*, 2018, pp. 5561–5568.
[9] G. Keren and M. Fridin, "Kindergarten social assistive robot (KindSAR) for children's geometric thinking and metacognitive development in preschool education: A pilot study," *Comput. Hum. Behav.*, vol. 35, pp. 400–412, Jun. 2014.

[10] M. C. Di Lieto *et al.*, "Educational robotics intervention on executive functions in preschool children: A pilot study," *Comput. Hum. Behav.*, vol. 71, pp. 16–23, Jun. 2017.
[11] J. Wan, S. Tang, H. Yan, D. Li, S. Wang, and A. Vasilakos, "Cloud robotics: Current status and open issues," *IEEE Access*, vol. 4, pp. 2797–2807, Jun. 2016.
[12] G. Hu, W. P. Tay, and Y. Wen, "Cloud robotics: Architecture, challenges and applications," *IEEE Netw.*, vol. 26, no. 3, pp. 21–28, May/Jun. 2012.
[13] Z. Du, L. He, Y. Chen, Y. Xiao, P. Gao, and T. Wang, "Robot cloud: Bridging the power of robotics and cloud computing," *Future Gener. Comput. Syst.*, vol. 74, pp. 337–348, Sep. 2017.
[14] B. Hu, H. Wang, P. Zhang, B. Ding, and H. Che, "Cloudroid: A cloud framework for transparent and QoS-aware robotic computation outsourcing," in *Proc. IEEE Int. Conf. Cloud Comput. (CLOUD)*, Jun. 2017, pp. 114–121.
[15] A. Chibani, Y. Amirat, S. Mohammed, E. Matson, N. Hagita, and M. Barreto, "Ubiquitous robotics: Recent challenges and future trends," *Robot. Auton. Syst.*, vol. 61, no. 11, pp. 1162–1172, 2013.
[16] C. Chakrabarti and G. F. Luger, "Artificial conversations for customer service chatter bots: Architecture, algorithms, and evaluation metrics," *Expert Syst. Appl.*, vol. 42, no. 20, pp. 6878–6897, 2015.
[17] Y. Guo, X. Hu, B. Hu, J. Cheng, M. Zhou, and R. Y. K. Kwok, "Mobile cyber physical systems: Current challenges and future networking applications," *IEEE Access*, vol. 6, pp. 12360–12368, 2018.
[18] M. Sneps-Sneppe and D. Namiot, "On context-aware proxy in mobile cloud computing for emergency services," in *Proc. 24th Int. Conf. Telecommun. (ICT)*, May 2017, pp. 1–5.
[19] W. Budiharto, A. D. Cahyani, P. C. B. Rumondor, and D. Suhartono, "EduRobot: Intelligent humanoid robot with natural interaction for education and entertainment," *Procedia Comput. Sci.*, vol. 116, pp. 564–570, Dec. 2017.
[20] J. Wan, S. Tang, Q. Hua, D. Li, C. Liu, and J. Lloret, "Context-aware cloud robotics for material handling in cognitive industrial Internet of Things," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2272–2281, Aug. 2018.
[21] X. Hu *et al.*, "Emotion-aware cognitive system in multi-channel cognitive radio ad hoc networks," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 180–187, Apr. 2018.
[22] Z. Ning *et al.*, "A cooperative quality-aware service access system for social Internet of vehicles," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2506–2517, Aug. 2018.
[23] X. Hu, X. Li, E. C.-H. Ngai, V. C. M. Leung, and P. Kruchten, "Multidimensional context-aware social network architecture for mobile crowdsensing," *IEEE Commun. Mag.*, vol. 52, no. 6, pp. 78–87, Jun. 2014.
[24] X. Hu *et al.*, "SAfeDJ: A crowd-cloud codesign approach to situation-aware music delivery for drivers," *ACM Trans. Multimedia Comput., Commun. Appl.*, vol. 12, no. 1s, p. 21, 2015.
[25] J. Zhang *et al.*, "Energy-latency tradeoff for energy-aware offloading in mobile edge computing networks," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2633–2645, Aug. 2018.
[26] Z. Liu, X. Zeng, W. Huang, J. Lin, X. Chen, and W. Guo, "Framework for context-aware computation offloading in mobile cloud computing," in *Proc. 15th Int. Symp. Parallel Distrib. Comput. (ISPDC)*, Jul. 2016, pp. 172–177.
[27] R. Roostaei and Z. Movahedi, "Mobility and context-aware offloading in mobile cloud computing," in *Proc. IEEE Int. Conf. Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People, Smart World Congr. (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)*, Jul. 2016, pp. 1144–1148.
[28] X. Wang, W. Wang, and Z. Jin, "Context-aware reinforcement learning-based mobile cloud computing for telemonitoring," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Inform. (BHI)*, Mar. 2018, pp. 426–429.

JIANBO ZHENG received the B.S. degree in management from the Shandong University of Science and Technology, China, in 2006, the M.S. degree in software engineer from Southeast University, China, in 2013, and the M.S. degree in computer science from the University of New Brunswick, Canada, in 2014. He is currently an Engineer with the Guangdong Provincial Key Laboratory of Robotics and Intelligent System, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. His research interests include cloud robotics, human–computer interaction, and data mining.



QIESHI ZHANG (S'07–M'14) received the B.E. degree in automation from the Xi'an University of Technology, China, in 2004, and the master's and Ph.D. degrees from Waseda University, Japan, in 2009 and 2014, respectively. From 2010 to 2012, he was a Research Fellow with the Japan Society for the Promotion of Science. From 2012 to 2016, he was a Research Assistant and a Research Associate with the Information, Production and Systems Research Center, Waseda University. From 2016 to 2018, he was an Assistant Professor with Shaanxi Normal University, China. He is currently a Senior Engineer with the Guangdong Provincial Key Laboratory of Robotics and Intelligent System, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China. He has authored or co-authored over 50 scientific articles in international journals and conferences. His current research interests are artificial intelligence, unmanned drive, image processing, computer vision, and machine learning. He serves as the technical/program committee member for over 50 conferences and over 100 times. He is a member of the ACM.

SHIHAO XU received the B.S. degree in communication engineering from Lanzhou University, Lanzhou, China, where he is currently pursuing the master's degree with the School of Information Science and Engineering. He is also a Guest Student with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. His major research interests are human–computer interactions and big data analyses.

HONG PENG received the Ph.D. degree from Lanzhou University, Lanzhou, China. From 2010 to 2011, he was as a Visiting Scholar with the Institute of Computer System, ETH Zurich, Switzerland. He is currently an Associate Professor with the School of Information Science and Engineering, Lanzhou University. He is in charge of three projects from National Natural Science Foundation of China, Central College Foundation Project of Lanzhou University and Youth Cross-Project of Lanzhou University. He is involved in the work of biosensors, biological signal processing, and emotional characteristics analysis. His research areas include bioinformation processing and ubiquitous affective computing.

QIN WU received the B.A. degree in art design from the Chengdu University of Information Technology, China, in 2012, and the M.F.A. degree in information art and design from Tsinghua University, Beijing, China, in 2016. She is currently a Lecturer with the School of Computer Science, Chengdu University of Information Technology, China. She is also a Research Assistant with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. Her research interests include human–computer interaction, user experience, and interactive design.