ISSN: 0974-5823 Vol. 6 No. 3 December, 2021

International Journal of Mechanical Engineering

Emotion Recognition system And Classifier for Facial Gestures Based on Machine Learning Techniques

Simrat Kaur Rakhraj
M. Tech. Scholar
School of Engineering and I.T.,
MATS University, Raipur

Apurv Verma
Assistant Professor
School of Engineering and I.T.,
MATS University, Raipur

Dr. Abhishek Badholia
Associate Professor
School of Engineering and I.T.,
MATS University, Raipur

Abstract: Here We present a unique methodology for detecting emotional responses from facial gestures that is aimed for educational processes. Our approach is developed on a multilayer deep neuronal network that identifies moods taken by a webcam. We utilize Russell's framework of core affect for our classification result, where any emotions could be classified into one of four dimensions. We collected information from several resources, normalised it, and used it to prepare the deep learning approach. We utilized VGG S network's fully linked layers, which were developed on manually labelled human face gestures. We put our application to the test by dividing the data by 80:20 and retraining the model. The cumulative efficiency of all identified moods in the test was 66%. On a normal personal computer featuring just a digicam, we have a working programme that can detect the person's behavioral position at roughly five frames per second. The emotional situation sensor will be embedded into an affective educational assistant framework and used to provide responses to a smart graphical instructional assistant.

Keywords: Neural-Networks, Facial-recognition, Deep-learning;

1 Introduction

The human face is the most powerful origin of behavioral detail, and face gestures are crucial in interpersonal interaction. Because of millions of years of evolution, we are capable of reading and comprehending facial expressions. We also respond to facial gestures, and few of such responses are unintentional [1]. Feelings are useful as learning suggestions since they inform the instructor about the participant's mental situation. This is especially essential in online education, where a completely automatic model can be tailored to the learner's expressive condition.

We introduce an unique emotion detection algorithms relying on deep neural networks for usage in academic settings [2]. Using a webcam, we gather photographs from the National Science Foundation's Collaborative Research: Multimodal Affective Pedagogical Agents for Various Kinds of Students. Learners and ConvNet identify facial gestures in instantaneously and classify them using Russell's classifier model and Figure -1, where it encompasses the large percentage of sensations a trainee can encounter throughout a learning event. We have put in the energy to cope through a variety of sceneries, observing viewpoints, and lights circumstances that might be experienced in real-world situations. We apply transmission knowledge to the VGG S network's fully linked layers, which were trained on manually labelled human face expressions. On a personal macbook, our program can indicate the participant's feelings condition at about 5-FPS, and the entire evaluation correctness of the recognised feelings were sixty-six Percentage [3]. The mental situation detectors will be integrated into an affective educational agent system, where it will provide as response to a smart illustrated teacher.

2 Literature Review

A feature sensor is used to evaluate a facial picture and a list of focus spots is acquired. Those focus spots were linked to the interest positions that represent a person's neutral expression. Deviations in the position of the interest point (x and y) are then investigated using a classifier to determine what emotion is being conveyed. We've picked six folks from around the world with seven different expressions (among which one is taken to be neutral) [4]. After that, the images are processed using Face++, an Application Programming Interface (API) that uses a CNN-based approach to provide cognitive services in the cloud. Face++ servers receive face photos as input and provide the x and y coordinates of eighty-three interest points for each face recognized. There are six main face emotions in each show: happiness, pain, curiosity, despair, rage, hatred, and disdain. The link among differences in the focus spots generated by Face ++ and the classification of the phrases saved in the servers is then verified. The fluctuation within every focus spot's position whenever the facial is neutral, – i.e., if there is no representation on the face, and its placement in each of the remaining 6 emotions were calculated. As a result, those variances are represented as an aggregate of the 6 people [5].

In instructional environments, encoding and interpreting feelings is very crucial. While direct instruction from a qualified, knowledgeable, and compassionate instructor is ideal, it is never usually achievable. Individuals have now been considering teaching without educators since the discovery of books, and more recently with technological breakthroughs, such as the use of simulations. Distance learning platforms and technologies have also advanced significantly. Although automating contains numerous benefits, including touching a large number of people or becoming accessible in places which face-to-face teaching isn't possible, it also poses new obstacles. One of them is the course's uniform appearance and feel [6]. Single pattern doesn't really suit every participant; transmission must be regulated; assignments might change according to the learner 's ability; as well as material must be adapted to the learners' specific requirements.

Affective Agents: Engaging instructional agents which have been demonstrated to be effective in boosting distant learning have addressed some of these problems. Graphical teaching entities perform a key role within them because they are simple to handle and their behavior may be defined using techniques prevalent in motion graphics, such as delivering appropriate gestures. According to Kim et al., educational agents with emotional capacities can increase learning by enhancing interactions between the student and the computer. Several methods have been built; for example, Lisetti and Nasoz used a combination of facial expression and physiological cues to detect the emotions of a learner [7]. D'Mello and Graesser presented Auto Tutor, demonstrating that learners exhibit a wide range of emotions when learning and that AutoTutor can be programmed to recognize and respond to these emotions.

Although, an effective educational agent must first identify and respond to learners' emotions. The interchange must premised on the model's actual input, educational agents should be empathic, and emotional exchanges with the learner should be provided. Expressions on the Face: While the earlier work discussed above produces excellent outcomes, it may not always be suitable in an educational setting. When engaging with instructional agents, speech is not always essential, and alternatives that involve the use of linked sensors may not be suitable for the student [9]. This remains face expressions recognition and evaluation as a viable alternative.

Several methods for detecting facial gestures have been presented. The FACS, for example, focuses on face parameterization, in which traits are identified and recorded as a feature vector that is employed to discover a certain mood. A vast class of algorithms, such as face reconstructions and others, aim to discover prominent facial characteristics using geometry-based methodologies [10]. Different sentiments, as well as their variants, have been investigated and classified, with some focusing on micro-expressions. Novel techniques employ machine learning techniques like SVM to automate feature recognition, but they have the same face detector sensitivities as the above-mentioned methodologies.

A facial tracking system, which must be competent of robust recognition of face and its characteristics under various light circumstances and for other models [11], is one of the fundamental components of these techniques. Existing approaches, on the other hand, frequently need meticulous calibration and identical illumination situations, and the standardization might not be transferable to other people. While this type of method are effective at detecting head position and orientation, they may miss minor variations in attitude that are necessary for sentiment recognition.

DL is a term that refers to Deep neural networks have been used to the field of emotion detection as a result of recent developments in deep learning. Several methods for detecting robust head rotation, facial characteristics, speech, and even emotions have been proposed [12]. EmoNets, for example, analyses both visual and audio streams concurrently to detect acted emotions in movies. This method improves on prior face detection work by CNN. Burket et al. created a deep learning network dubbed DeXpression for emotion identification from videos, which motivated our study. They employ the Cohn-Kanade database and the MMI Facial Expressions in specific.

3 Classification of Emotions

Most emotion recognition software divide photos of facial gestures into seven categories: anger, disgust, worry, joyful, sorrow, shock, and neutral. In context of students' emotions, such classification is overly thorough; for example, when students are attending audiovisual lectures in view of a screen, the high number of feelings is not appropriate in all instances [13]. As a result, we employ a classification of emotions that are connected to and utilized in academic learning. We employ Russell's concept of core affect, in which every emotion may be classified along 2D. This approach covers a wide variety of feelings and is appropriate for deep learning application [14].

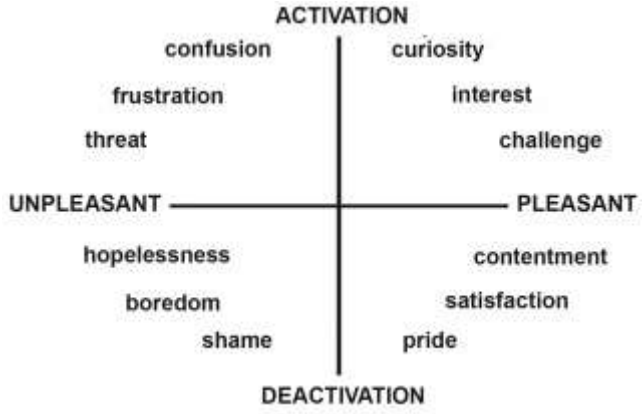

Figure 1: Emotional mappings four-quadrant model

The Russel's two major axes split the emotional spectrum onto four corners:

1. The top-left quadrant contains affective emotions associated with education, such as perplexity or frustration.
2. Curiosity and Interest are found at top-right corners;
3. Happiness and Fulfilment are found in the down-right corners;
4. Hopelessness and boredom are found in the down-left corners.

The majority of extant picture databases categorize photographs of facial gestures into the 7 distinct sentiments listed above [15]. By categorizing the photos utilizing the follow ing mappings, we change the sets of data as per Russell's 4-quadrants classification algorithm:

pleasant − active ⇐ happy, surprised,

unpleasant − active ⇐ angry, fear, disgust,

pleasant − inactive ⇐ neutral, and

unpleasant − inactive ⇐ sad.

This grouping then assigns a unique label denoted by L to each image as:

$$L \in \{\text{active} - \text{pleasant}, \text{active} - \text{unpleasant}, (1)$$

$$\text{inactive} - \text{pleasant}, \text{inactive} - \text{unpleasant}\}.$$

Methods

There are a number of databases of classified (labelled) facial gestures with recognized facial attributes. We utilized pictures from Various Databases declared in the table:

Table 1. Databases that are utilized to develop the deep neural network.

Input database	No. of images	sad	happy	neutral	surprise	fear	anger	disgust
CK+ [27]	636	28	69	327	83	25	45	**59**
JAFFE [38]	213	31	31	30	30	32	30	**29**
MUG [1]	401	48	87	25	66	47	57	**71**
ISED [23]	478	48	227	0	73	0	0	**80**
RaFD [32]	7,035	1,005	1,005	1,005	1,005	1,005	1,005	**1,005**
Oulu [67]	5,760	480	480	2,880	480	480	480	**480**
AffectNet [41]	28,7401	25,959	134,915	75,374	14,590	6,878	25,382	**4,303**
CMU-PIE [21]	**551,700**	**0**	**74,700**	**355,200**	**60,900**	**0**	**0**	**60,900**

Table 1 shows the amount of photos in every dataset as well as the categorization of each collection. Figure 1 depicts the data distributions by expression, database, and the percentile distribution of every emotion in the collection. We had 863,624 photos in all, with 52 percent neutral looks, 24 percent joyful faces, 3 percent sad faces, 8 percent disgust, 3 percent rage, 1 percent fear, and 9 percent astonishment [19].

There were 288,741 active-pleasant photos (12%), 102,393 active-unpleasant images (34%), 434,841 inactive-pleasant images (51%), and 27,599 inactive-unpleasant images (3%) [20].

Certain photographs may be classified by only one person, but others may be classified by several persons, adding to the confusion. Furthermore, few records are in colour, while others are not. Whereas it would be superlative if all databases had uniform coverage of face expressions, the databases are uneven in terms of image quality and coverage of facial emotions.

Furthermore, although certain phrases are easily classified, others might be categorized as mixed and association to numerous groups [22]. In this scenario, we eliminated picture from the trials or assigned it to a single set. Surprisingly, the furthermost problematic expression to describe is unbiased, as it lacks emotional energy and can be readily misunderstood. This phrase is perhaps the largest covered in the dataset, which might increase recognition if properly prepared [23].

Training

The network was trained using photos from the datasets presented in Section 4.1. To improve our program's accuracy in actual circumstances, we applied data amplification techniques such as Gaussian blur and changes in contrast, lighting, and subject location on the original photos from each dataset [26]. The input photos were preprocessed using OpenCV's Haar-Cascade filter, which crops the image to include only the face and no significant backgrounds. As a result, instructional periods are reduced [27, 28].

To get a balancing training data, we'd want to have a same quantity of photos for every identifier from the category in Eq (1). [29, 30] Finally, the size of the training set was determined by the fewest photos. We trained using 68,012 photos, with a batch size of 15 pictures, 80,000 rounds, and an averaged correctness of 0.63 with 10,000 epochs [31]. The training took around 70 minutes on a device outfitted with an Intel Xeon(R) W-2145 CPU operating at 3.7 GHz, 32 GB of RAM, and an NVidia RTX2080 GPU [32, 33].

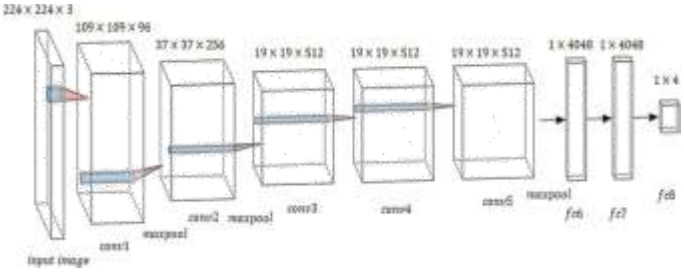

Figure 2: Our approach employs DNN architecture.

Results

We divided the material into 2 subgroups irregular intervals. We trained on 80% of the photographs, assessed on 20%, and then re-performed the procedure 3-times with a different random distribution of the instructions every round.

Table 2. Our testing's averaged and standard deviations.

	Pleasant Active	**Pleasant Inactive**	**Unpleasant Active**	**Unpleasant Inactive**
Pleasant Active	(71, 1.5)	(22.4, 4.2)	(2.1, 0.0)	(6.4, 2.1)
Pleasant Inactive	(2.4, 1.1)	(87.4, 0.5)	(4.2, 0.8)	(5.9, 1.5)
Unpleasant Active	(1.6, 0.4)	(41.8, 0.8)	(45, 1.6)	(8.4, 5.7)
Unpleasant Inactive	(6.0, 0.9)	(11.0, 3.8)	(9.3, 2.9)	(63.0, 7.1)

The average and standard deviations of the confusion matrices from the three runs of our studies are shown in Table 2, and the error matrix of the individual runs are shown in Fig. 3. This is an expected conclusion since the lowest portion of the Russel's diagram contains passive expressions, which are more difficult to identify in general. We attained a 66 percent total accuracy.

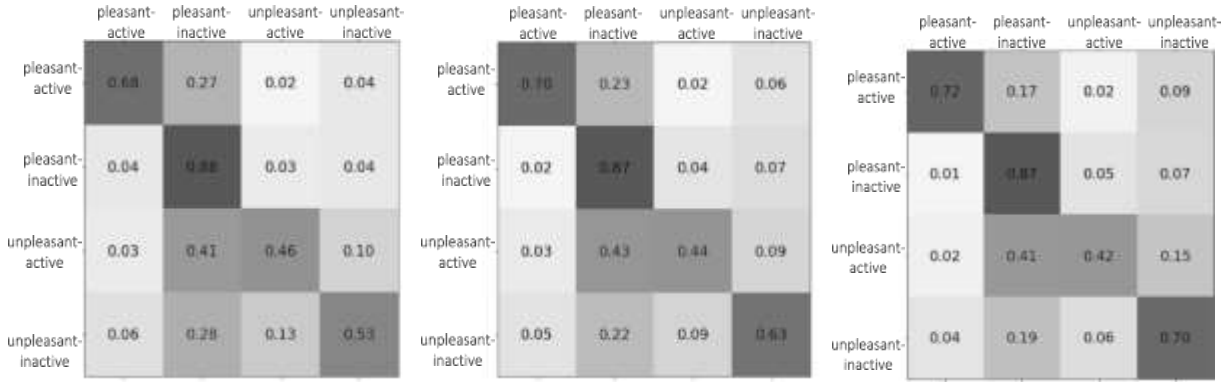

Figure 3: Regarding overall outcomes of this study, we used standardized confusion matrix.

The trained DNN was retrieved and utilized in actual period to classify emotional gestures into the 4 categories of Russell's diagram. We employed a personal laptop with just a digicam having a Quality of 1920x1080 px and a 2.4ghz Band Intel Core i5 CPU. For monitoring the incoming images via the webcam & recognize faces, we employed the Caffe environment on Windows 10 and OpenCV. The background was cut out and the front is transmitted to our skilled system. The image was classified by the NN, and the result was transmitted back to the program, which showed it as a label on the monitor. Figure 5 depicts many examples of real-time facial emotion recognition employing our technique.

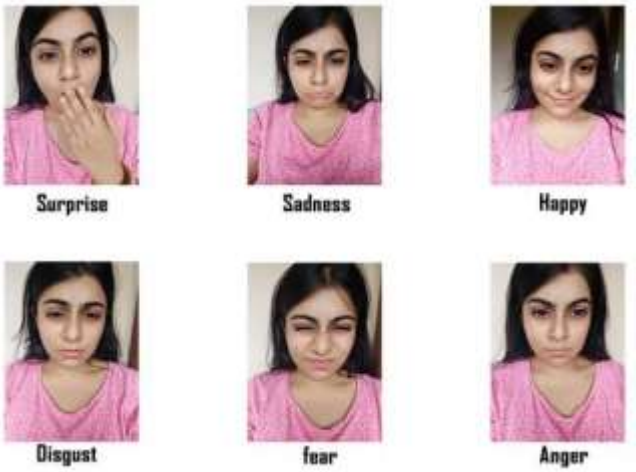

Figure 5: Legitimate expressions recognition demonstrations employing our technique.

Conclusions

The objectives of research work are to create a proper face expression identification system that identifies and classifies individual sentiments, with the goal of employing it as a web-based learning classifier. Our sensor indicates a likelihood of an emotions assigned to one of 4 categories of Russel's diagram as a result of this condition.

The detection technique will be integrated into a framework of affective educational entities that will interact to the children's identified sentiments utilizing various sorts of sentimental understanding in the upcoming years. Our tests suggest that total test correctness is adequate for realistic application, and we anticipate that the whole program would improve understanding.

We executed an initial user research where ten persons were instructed to produce specific facial expressions and the recognition was confirmed. This technique, although, failed to deliver satisfying outcomes since we are unable to authenticate that individuals were all in the appropriate mental situation and therefore that their emotions are real - several individuals began to giggle whenever the algorithm recognized emotions they weren't really expected. Sentimental condition is a difficult thing to understand. Even for genuine performers, some of the looks were difficult to accomplish since happy individuals cannot push them self to portray serious faces.

References

[1]. Zhou W., Cheng J., Lei X., Benes B., Adamo N. (2020) Deep Learning-Based Emotion Recognition from Real-Time Videos. In: Kurosu M. (eds) Human-Computer Interaction. Multimodal and Natural Interaction. HCII 2020. Lecture Notes in Computer Science, vol 12182. Springer, Cham. https://doi.org/10.1007/978-3-030-49062-1_22.

[2]. J. M. López Gil and N. Garay Vitoria, "Photogram Classification-Based Emotion Recognition," in IEEE Access, vol. 9, pp. 136974-136984, 2021, doi: 10.1109/ACCESS.2021.3117253.

[3]. Aifanti, N., Papachristou, C., Delopoulos, A.: The MUG facial expression database. In: 11th International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS 2010, pp. 1–4. IEEE (2010).

[4]. Bettadapura, V.: Face expression recognition and analysis: the state of the art. arXiv preprint arXiv:1203.6722 (2012).

[5]. Borth, D., Chen, T., Ji, R., Chang, S.F.: SentiBank: large-scale ontology and clas- sifiers for detecting sentiment and emotions in visual content. In: Proceedings of the 21st ACM International Conference on Multimedia, pp. 459–460 (2013).

[6]. Burkert, P., Trier, F., Afzal, M.Z., Dengel, A., Liwicki, M.: DeXpression: deep convolutional neural network for expression recognition. arXiv preprint arXiv:1509.05371 (2015).

[7]. Castellano, G., et al.: Towards empathic virtual and robotic tutors. In: Lane, H.C., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013. LNCS (LNAI), vol. 7926, pp. 733–736. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39112- 5 100.

[8]. Fayek, H.M., Lech, M., Cavedon, L.: Evaluating deep learning architectures for Speech Emotion Recognition. Neural Netw. 92, 60–68 (2017).

[9]. Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-PIE. Image Vis. Comput. 28(5), 807–813 (2010).

[10]. Gunawardena, C.N., McIsaac, M.S.: Distance education. In: Handbook of Research on Educational Communications and Technology, pp. 361–401. Routledge (2013).

[11]. Happy, S., Patnaik, P., Routray, A., Guha, R.: The indian spontaneous expres- sion database for emotion recognition. IEEE Trans. Affect. Comput. 8(1), 131–142 (2015).

[12]. Cheng, J., Zhou, W., Lei, X., Adamo, N., Benes, B.: The effects of body gestures and gender on viewer's perception of animated pedagogical agent's emotions. In: Kurosu, M. (ed.) HCII 2020. LNCS, vol. 12182, pp. 169–186. Springer, Cham (2020).

[13]. Kahou, S.E., et al.: EmoNets: multimodal deep learning approaches for emotion recognition in video. J. Multimodal User Interfaces 10(2), 99–111 (2016). https://doi.org/10.1007/s12193-015-0195-2.

[14]. Kanade, T., Cohn, J.F., Tian, Y.: Comprehensive database for facial expression analysis. In: IEEE International Conference on Automatic Face and Gesture Recognition, pp. 46–53. IEEE (2000).

[15]. Kim, S., Georgiou, P.G., Lee, S., Narayanan, S.: Real-time emotion detection sys- tem using speech: multi-modal fusion of different timescale features. In: 2007 IEEE 9th Workshop on Multimedia Signal Processing, pp. 48–51. IEEE (2007).

[16]. Kim, Y., Baylor, A.L.: Pedagogical agents as social models to influence learner attitudes. Educ. Technol. 47(1), 23–28 (2007).

[17]. Kim, Y., Baylor, A.L., Shen, E.: Pedagogical agents as learning companions: the impact of agent emotion and gender. J. Comput. Assist. Learn. 23(3), 220–234 (2007).

[18]. Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H., Hawk, S.T., Van Knip- penberg, A.: Presentation and validation of the Radboud Faces Database. Cogn. Emot. 24(8), 1377–1388 (2010).

[19]. Le, Q.V., Zou, W.Y., Yeung, S.Y., Ng, A.Y.: Learning hierarchical invariant spatio- temporal features for action recognition with independent subspace analysis. In: CVPR 2011, pp. 3361–3368. IEEE (2011).

[20]. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature 521(7553), 436–444 (2015).

[21]. Levi, G., Hassner, T.: Emotion recognition in the wild via convolutional neural networks and mapped binary patterns. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, pp. 503–510 (2015).

[22]. Lyons, M., Kamachi, M., Gyoba, J.: Japanese Female Facial Expression (JAFFE) Database, July 2017. https://figshare.com/articles/jaffe desc pdf/5245003.

[23]. Martha, A.S.D., Santoso, H.B.: The design and impact of the pedagogical agent: a systematic literature review. J. Educ. Online 16(1), n1 (2019).

[24]. Mollahosseini, A., Hasani, B., Mahoor, M.H.: AffectNet: a database for facial expression, valence, and arousal computing in the wild. IEEE Trans. Affect. Com- put. 10(1), 18–31 (2017).

[25]. Morency, L.P., et al.: SimSensei demonstration: a perceptive virtual human inter- viewer for healthcare applications. In: Twenty-Ninth AAAI Conference on Artificial Intelligence (2015).

[26]. Neri, L., et al.: Visuo-haptic simulations to improve students' understanding of friction concepts. In: IEEE Frontiers in Education, pp. 1–6. IEEE (2018).

[27]. Ng, H.W., Nguyen, V.D., Vonikakis, V., Winkler, S.: Deep learning for emotion recognition on small datasets using transfer learning. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, pp. 443–449 (2015).

[28]. Russakovsky, O., et al.: ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. 115(3), 211–252 (2015). https://doi.org/10.1007/s11263-015-0816-y.

[29]. Schroeder, N.L., Adesope, O.O., Gilbert, R.B.: How effective are pedagogical agents for learning? A meta-analytic review. J. Educ. Comput. Res. 49(1), 1–39 (2013).

[30]. Tie, Y., Guan, L.: A deformable 3-D facial expression model for dynamic human emotional state recognition. IEEE Trans. Circ. Syst. Video Technol. 23(1), 142–157 (2012).

[31]. Yang, S., Luo, P., Loy, C.C., Tang, X.: From facial parts responses to face detection: a deep learning approach. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3676–3684 (2015).

[32]. Yuksel, T., et al.: Visuohaptic experiments: exploring the effects of visual and haptic feedback on students' learning of friction concepts. Comput. Appl. Eng. Educ. 27(6), 1376–1401 (2019).

[33]. Zhou, L., Mohammed, A.S., Zhang, D.: Mobile personal information management agent: supporting natural language interface and application integration. Inf. Process. Manag. 48(1), 23–31 (2012).