Facial Expressions based Error Detection for Smart Environment Using Deep Learning

[1]Yacine Yaddaden, [2]Mehdi Adda, [1]Abdenour Bouzouane, [1]Sebastien Gaboury & [3]Bruno Bouchard
[1]*Université du Québec à Chicoutimi (UQAC)*
[2]*Université du Québec à Rimouski (UQAR)*
[3]*IEEE Senior, ACM Professional Specialist in Artificial Intelligence*
Email: {yacine.yaddaden1, abdenour_bouzouane, sebastien_gaboury}@uqac.ca, mehdi_adda@uqar.ca

***Abstract*—Creating innovative applications of ambient intelligence for smart environments became a real challenge. Indeed, the existing systems are considered as being obtrusive. Therefore, the ongoing trends consist in finding new methods which are less intrusive while being accurate and computer efficient.To that end, we introduce an ingenious way to assist the elderly and people with cognitive impairment in their daily life using emotions through facial expressions. Our hypothesis assumes the correlation between expressed emotions and the resulting errors in daily activities. In order to verify it, we introduce an experimentation protocol with neurotypical subjects. In this paper, we mainly focus on designing an approach for facial expression recognition using Convolutional Neural Network. The carried experiments with benchmark datasets provided promising results. In fact, the proposed approach outperforms the existing methods in terms of accuracy. Moreover, our approach has been implemented in an embedded system with limited resources. The main objective of this paper is to provide an effective approach recognizing facial expressions that will be used for our further experimentations.**

***Keywords*-Ambient Assisted Living; Ambient Intelligence; Data Mining; Machine Learning; Convolutional Neural Network; Facial Expressions; Emotion Recognition.**

I. Introduction

In the last decades, the world went through critical demographic changes. In fact, the world has seen an important increase of ageing populations. Ageing people are in constant need of assistance, particularly those with cognitive impairments. Despite the increase of healthcare facilities, the elderly are interested in technologies allowing them to stay at home. The motivations consist in the high costs of such services and the desire of ageing people to keep their usual environment and their habits. Therefore, different fields such as AmI (Ambient Intelligence) and AAL (Ambient Assisted Living) have emerged and various solutions have been introduced to ensure the well-being of people with cognitive impairment while staying at home.

The most relevant illustration of AAL consists in smart homes. They might be described as common living places equipped with sensors to collect environmental data and actuators to perform adequate actions. In order to provide suitable assistance, it is important to recognize and analyze the activities of the elderly. However, a crucial part of assistance consists in error detection. It is common sense to think that the elderly need help only when they make mistakes or they do not remember how to perform a task. Information provided through activity recognition is not enough to detect errors and suffers from high false positive rate. It may appear to the elderly annoying and intrusive. Usually, in order to detect errors, existing approaches exploit the sequence order of the different steps in a specific task. If an anomaly is detected during that activity, it is automatically reported as an error and an aid response is triggered. However, it is not necessarily a mistake; the user might intentionally interrupt his task or he may change his habit as: putting or not sugar in his coffee.

The final objective of our work is to introduce a novel approach to detect errors in ADLs (Activities of Daily Living) using information provided from the user himself. The proposed approach exploits facial expressions as data collected from the user's face. We believe that when people make mistakes, they express specific emotions through facial expressions. Our work aims to verify the presence of a correlation between making mistakes during ADLs and expressed emotions. Before that, the first step consists in developing a robust and accurate approach for facial expressions recognition. The targeted environment where the proposed approach will be used is limited in terms of hardware resources which means that our developed approach must take into account these constraints. The perfect candidate is the CNN (Convolutional Neural Network) which is a popular technique of DL (Deep Learning) [1]. CNN has allowed to reach the highest rates in terms of recognition for different applications, but it has some flaws such as the dependency in high performance hardware. Our challenge is to develop an effective CNN architecture that will be accurate and adapted to limited resources environments.

A question might be asked about the motivation behind the choice of facial expressions as modality for recognizing emotions. In fact, various other modalities might be used for recognizing emotions such as: physiological signals, speech, voice, body gesture. However, according to Mehrabian [2], facial expressions contribute by 55% to the message effect

978-1-5386-0435-9/17/$31.00 ©2017 IEEE

while 7% and 38% are respectively attributed to the verbal and vocal part. Numerous works have been conducted in the field of facial expressions. The most common and popular studies are attributed to Paul Ekman [3] who described the expressed emotions following six basic ones: happiness, fear, anger, surprise, disgust and sadness.

The present work might be described by two distinct but dependable parts. The first one consists in the development of an approach to recognize emotions through facial expressions using a DL technique. The challenge is to optimize the CNN using a specific architecture in order to achieve two objectives at the same time. Indeed, the proposed architecture must be able to be implemented in a constrained environment in terms of available computational resources and being sufficiently accurate with a high recognition rate. In order to verify the efficiency of the approach, it has been tested with benchmark datasets in a well-known embedded system. Finally, we propose an experimental protocol to verify the correlation between errors in ADLs and facial expressions.

The rest of the paper is organized as follows: In section II, we present various existing work related to facial expressions recognition using classical and DL techniques, fundamentals of error detection in ADLs and the use emotion for AAL and AmI. In section III, we describe each component of the proposed approach. Then in section IV, we introduce the different steps of the experimental protocol. In section V, we present the results of our experimentations using two common facial expressions datasets. Finally, in section VI, we discuss our future work and perspectives.

II. Related Works

A. Fundamentals

A basic system of emotion recognition from facial expressions has the same building blocks as a classical pattern recognition system. The system might be designed by different ways depending on the input signal: *static* or *dynamic*. In both cases, *feature extraction* provides a better representation of the input signal using specific types of characteristics. In the field of facial expressions recognition, three different kinds of features might be used: *geometric-based* exploit the facial fiducial points, *appearance-based* rely on the texture of the entire image and *hybrid-based* features combine the two previous ones. Considering the noisy and redundant aspect of features, *selecting features* chooses and holds the most discriminant and representative ones. *Classification* allows to construct a model using prepared training data. The model aims to recognize the facial expression from an unlabeled input signal.

B. Classical Techniques

To illustrate the fundamentals that we have introduced, we present several existing approaches based on *shallow* architectures. For the geometric-based features, Hsu et al. [4] and Yaddaden et al. [5], introduced approaches using extracted facial fiducial points as characteristics and combined them with two distinct classifiers: SVM (Support Vector Machine) and KNN (K-Nearest Neighbors), respectively. Other approaches exploit appearance-based characteristics such as Zhong et al. [6]. They proposed to export LBP (Local Binary Pattern) as a descriptor combined with an SVM classifier. For the same type of features, Rao et al. [7] proposed to use SURF (Speeded-Up Robust Features) as descriptors combined with a Gentle Adaboost with logistic regression. Several authors proposed to merge the two previous types of characteristics to improve accuracy. Youssif et Asker [8] used facial characteristics points with coefficients collected when applying Canny Edge Detector. For classification, they exploit an AAN (Artificial Neural Network) classifier. Although classical approaches provide interesting results, they still suffer from different drawbacks such as the need of optimizing the three different blocks (feature extraction, feature selection & classification). In fact, a system might provide highest accuracy with a specific dataset, but it will not be the same with a different one. Another flaw consists in the *hand-engineering* features. The characteristics are manually chosen by the user who conceives the system.

C. Deep Learning Techniques

In order to address the different issues of the presented approaches, some researchers have recently used *deep* architectures related to DL. Among these techniques: RNN (Recurrent Neural Network), DBN (Deep Belief Network), SAE (Stacked Auto Encoder), CNN and other related methods. The main advantage of DL approaches is the presence of a unique block. It aims to perform the different steps of a classical pattern recognition system. The autonomous feature extraction without human intervention might be considered as another advantage. These benefits motivated researchers of different areas to exploit DL techniques particularly in computer vision. In the field of facial expression recognition, various works have been conducted using DL methods. Lv et al. [9] proposed an approach using DBN for feature extraction and SAE for classification. Li et al. [10] introduced the use of a CNN architecture combined with pre-processing steps in order to improve the accuracy. Mayya et al. [11] used deeper architectures called DCNN (Deep CNN) which contain more components and layers that allow to perform more complex processing. Shin et al. [12] proposed to compare common CNN architectures such as Caffe-ImageNet [13] and used them for facial expression recognition. They also added different techniques for image enhancement as pre-processing steps before feeding the CNN. These approaches enabled to break through various challenges, but they suffer from certain limitations such as the dependency on high performance computing platforms. *Overfitting* is another critical problem which is due to the small amount of data. It might be noticed when the model

provides high recognition rates with training set and lower with validation set. Recent works have allowed to overcome this issue using two distinct techniques: *dropout* which was introduced by Srivastava et al. [14] and *data augmentation*.

D. Error Detection

Before going deeper and introducing the proposed approach, we present fundamentals about various tools used to assist people with cognitive impairments. We also present several existing systems that exploit emotions in the context of AAL and AmI. Bouchard et al. [15] have introduced a keyhole plan recognition model aiming to recognize the individual's behavior when performing ADLs. They have combined activity recognition and error detection to provide assistance in adequate situations. The authors have targeted people with Alzheimers disease when designing the error detection part of the system. They have used the works of Baum et al. [16] as prerequisites. Indeed, they have estimated that people suffering from Alzheimers disease might make six categories of errors: (1) *Initiation* errors happen when the individual is unable to begin an activity and might be explained by a loss of memory; (2) *Organization*, the different steps of the activity are performed in an inadequate order; (3) *Realization*, happen when the individual is distracted and add unnecessary steps to the main task; (4) *Sequence* based errors are noticed when the individual omits to execute a previous dependent step for the current one; (5) *Judgment* are directly related to the safety of the individual and mistakes that might harm him; (6) *Completion*, happen when the individual is unable to finish his current activity. We notice that there are numerous types of errors which might not be easy to detect using common existing techniques. The initiation errors as an example, happen before the individual begins the activity and there is no information about it. There is no way to know whether the individual wanted to perform an activity because it is all about his intentions.

Various researchers have proposed methods for providing help in specific ADLs such as Mihailidis et al. [17] proposed a system called COACH helping people with cognitive impairment in the task of hand washing. Peters et al. [18] proposed a system called TEBRA providing assistance when brushing teeth. Jean-Baptiste et al. [19] introduced a framework whose purpose is to help people with disabilities in the task of tea making. Even if these approaches might be effective in terms of assistance, the interaction model still far from the human being one.

Due to the importance of the interaction quality in an assistance process, several approaches have been proposed to improve the HCI (Human Computer Interaction). Emotions and particularly through facial expressions are considered as promising candidates. Therefore, various assistant robots have been equipped with emotion recognition. Broekens et al. [20] presented a review of several robots with cognitive capabilities used in smart homes as companions such as: Paro, iCat and Pearl. De Carolis et al. [21] proposed a robot which combines social interaction and assistance. The robot uses emotion recognition based on two different modalities: facial expression and speech. The conducted experiments have proved that emotions enhance acceptance of technology by the elderly.

III. Proposed Approach

In this section, we describe the different stages of our approach. The first one is the *pre-processing stage*. It allows the preparation and enhancement of the input image quality. In the second stage, *classification stage*, facial expressions are recognized.

A. Pre-processing stage

Our approach uses CNN which means that the features are automatically extracted. CNN is sensitive to the quality of the input images in terms of *orientation* and *illumination*. For this reason, different pre-processing steps have been introduced.

Before performing any pre-processing, we extract the Facial Characteristics Points (FCP). We exploited the method introduced by Kazemi and Sullivan [22] which is fast, accurate and efficient. Their method extracts 68 FCPs in a "millisecond". The first pre-processing step is *spatial normalization* and consists in correcting the alignment of the face. We only exploit two points of the whole detected FCPs; they are related to the eyes centers. In order to estimate the correction angle, a line is drawn linking the two eyes centers. It is compared to a perfect horizontal line to get the value of the correction angle. Based on it, the whole image is rotated. The next step consists in *face extraction* and we have used the bordered FCPs as reference to extract the face region.

The previous operations enabled to rectify the orientation and extract the zone of interest from the input image. However, we still need to solve other problems related to brightness and illumination. To that end, we used a common method which has been widely exploited to process images due to its efficiency and simplicity. It is *linear intensity normalization* and it works on a pixel level by applying the following formula:

$$I_{new} = (I - Min)\frac{Max_{new} - Min_{new}}{Max - Min} + Min_{new} \quad (1)$$

The I_{new} represents the new treated image in terms of texture while I represents the original one. The values of Max_{new} and Min_{new} are respectively set to 255 and 0 while Max and Min represent respectively the highest and lowest value of the pixel intensity in the input image. We also used another image enhancement technique which is widely used in the field of computer vision: *histogram equalization*. It works on the image histogram by adjusting it in order to have a uniform distribution of the pixel's intensity

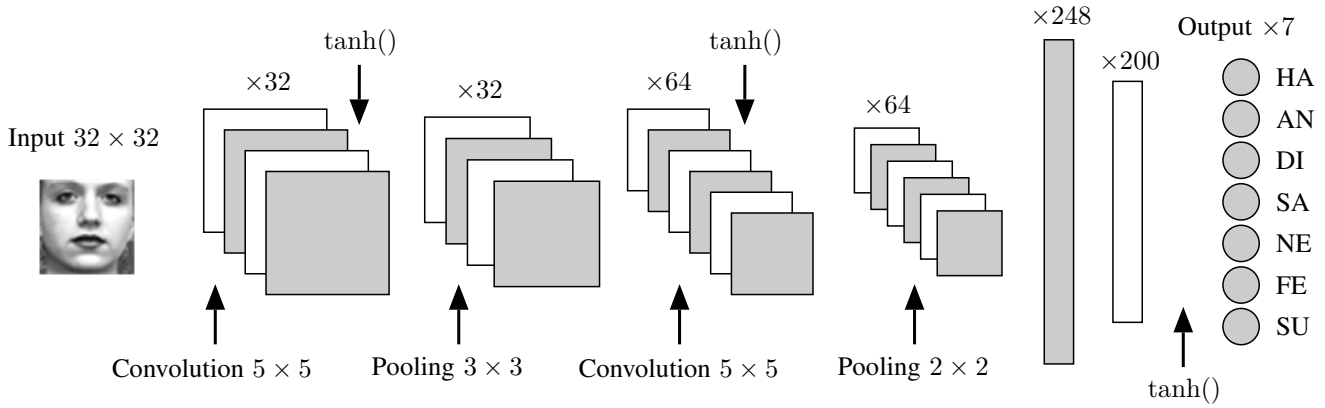

Figure 1. The CNN Architecture

value. Thus, the lower contrast pixel will be increased and vice versa.

The final step of the pre-processing stage is the image resizing, all images in the dataset must have the same size in order to be used in the *classification stage*. This parameter plays a crucial role in next stage because choosing a small size might induce an insufficiency of information which can be exploited to perform classification. In contrast, if the image is too big, it may increase computing time and resource consumption. In the context of constrained environments, we have chosen to use the smallest image size which is equal to 32×32.

B. Classification stage

For this stage, we used the CNN classifier which is a supervised machine learning technique. Therefore, it will perform emotion recognition from unlabeled face images after being trained with a certain facial expression dataset. We have been motivated to use it based on the recent promising results [9] [11]. Moreover, the CNN remedies to the hand-engineering feature extraction issue. As stated earlier, our main goal is to find a way to adapt the CNN architecture in order to embed it in a constrained environment without losing its accuracy.

We propose to use the CNN architecture as shown in Figure 1. It contains different layers and each one plays its own role. We have used only two *convolutional* layers whose role is to extract the relevant features from the input image. By analogy with classical methods, convolution layer plays the role of both *feature extraction* and *feature selection*. This operation might be described by the following formula:

$$S(i,j) = \sum_m \sum_n I(i-m, j-n) K(m,n) \qquad (2)$$

Where S represents the *feature map* and K the used *kernel* for the *convolution* operation. The next layer is called *detector* and basically, it consists in applying a *nonlinear activation function*. Its role is to ensure that the representation in the input space is mapped to a different space in the output. There are different nonlinear activation functions but for our application, the *hyperbolic tangent* showed the best performances and enabled to achieve the highest recognition rate. The *pooling* is another type of layers. It performs the replacement of a group of close pixels in the feature map by a summary statistic. We have used the *max-pooling* which consists in finding the highest value of a rectangle of pixels. It is described by the following formula:

$$y_{i,j,k} = \max_{pq}(x_{i,j+p,k+q}) \qquad (3)$$

$y_(i,j,k)$ is the output with i represents the i^{th} *feature map* and j,k are the co-ordinates. The parameters p,q represent the neighborhood of the *max-pooling*. The aim of this operation is to make the output representation become *invariant* to small translations of the input. The later layer is called a *fully-connected hidden layer* which map the different features maps into a one-dimensional vector. Based on this vector, the classifier will generate seven different responses (the six basic emotions and the neutral state).

Previous works have shown that Deep Learning approaches suffer from *overfitting* and the CNN is not an exception. In order to remediate to this issue, we have used two different approaches. The first one is adding *dropout* layers to the architecture. It has proved its efficiency in reducing the overfitting effect. The second solution consists in increasing the size of the dataset by generating synthetic images from an original one with four different rotation angles: $-6°, -3°, 3°, 6°$.

IV. Experimental Process

In the previous section, we have described in detail our approach to recognize emotions from facial expressions using CNN. As we want to use this system in the context of elderly assistance for error detection, we introduce a multi-step experimental protocol. The main purpose is to verify the correlation between facial expressions and errors when

performing a specific task. Another objective is to adapt the previously described approach of emotion recognition to error detection systems. By adding activity recognition techniques, it will form a complete and adapted system of assistance for the elderly and people with cognitive impairment.

In order to conceive an experimentation that reflects the real-world tasks, we have chosen to simulate an ADL that is preparing tea. Initially, the targeted population consists in people with cognitive disabilities but for the first phase, we work with neurotypical people without any cognitive impairment. The aim of this phase is to validate the experimental protocol before testing it with people suffering from cognitive disabilities. However, these people react differently than neurotypical ones when making errors in the context of an ADL. For this reason, we intentionally introduce anomalies in the process of tea preparation in order to observe the reactions of the subject through his facial expressions. The whole experiment is described according to four different stages and each one is detailed as follows.

A. Facial expression recognition

Before performing the experiments with real subjects, the first stage consists in developing an approach to recognize facial expressions. It is the main subject of this paper. We introduce a DL based approach validated using benchmark datasets. We have been motivated by the CNN capabilities in terms of learning and classification of images. There is also the autonomous extraction and selection of characteristics. However, we brought optimizations in order to make the CNN architecture implementable in embedded systems with limited hardware resources. The obtained results are promising and allow us to go further in the experimentations.

B. Collecting data

The second stage aims to collect the raw data that will be used to train the system for error detection and activity recognition. The experimentation is planned to be done in the LIARA (Laboratoire dIntelligence Ambiante pour la Reconnaissance dActivit) laboratory. It consists in a smart home equipped with various sensors and actuators. In the context of this experimentation, the used objects will have RFID (Radio Frequency IDentification) tags in order to track them if they are currently used during the ADL. All the needed objects will be placed on a table in order to avoid any movement of the subject. We will need two different cameras, the first one will be used to record the subject's face and the other one to record the whole scene. Three different samples will be recorded with each neurotypical subject: (1) A paper with all the different steps will be placed on the steps written on in order to guide the subject. No anomalies will be inserted during the process of tea making. (2) It will be exactly the same as the first sample with two differences. They consist in removing the papers with all the steps and adding a source of disturbance such as music or movies dialogue. (3) the third and last sample will be done by adding some anomalies during the process such as: empty the sugar or the tea box.

At the end of the experiment, we obtain three types of records or data. The first one is binary and received from the RFID tags placed on the various used objects. There are also video records of the whole scene monitoring every action of the subject. The last one consists in the video records of the subjects's face. It will be used to extract facial expressions in the context of emotion analysis.

C. Analyzing & extracting informtaion

After collecting all the needed raw data from the previous stage, the next one consists in the analysis and extraction of the useful information from the previous raw data. In the beginning will be used segmentation and clustering techniques in order to isolate the variations in facial expressions. As result, only small samples that represents the facial expression variation will be extracted from the whole facial video records and saved as images. Using the video records of the whole scene, we will be able to label the previously extracted samples. Among them, there will be those associated with making mistakes. The last type of data provided by RFID tags in the different objects will not be used as the aim of this study is error detection using facial expressions. However, it will be useful to design a real and complete system of assistance for the elderly and people with cognitive disabilities.

D. Building the model

The last stage of our experiment consists in training the previous system to detect errors during ADLs. The dataset is provided from the previous stage by labeling the different images corresponding to making errors. In this stage, we will also check the efficiency of DL based system in terms of assistance by recognizing errors using facial expressions.

V. Results & Discussion

A. Facial expression datasets

The experimental protocol which allows us to verify the correlation between facial expressions and errors made during ADLs has not be done yet, it has only been planned and described in the previous section. However, in order to validate our proposed approach of facial expression recognition, we have used common and benchmark datasets which have been exploited to measure the accuracy of existing approaches. In the following, we will describe two different facial expression datasets.

The CK+ (Cohn-Kanade Extended) [23] is a popular facial expression dataset. It includes 593 image sequences from 123 different subjects. This dataset contains female and male subjects of different ages. The images are grayscale. The second used dataset is the KDEF (Karolinska Directed

Table I
COMPARISON WITH OTHER EXISTING APPROACHES (CK+)

Methods	Classifiers	Recognition Rate
Hsu et al. [4]	SVM	92.90%
Zhong et al. [6]		89.89%
Youssif et Asker [8]	ANN	93.00%
Lv et al. [9]	DBN + SAE	91.11%
Mayya et al. [11]	DCNN	96.02%
Li et al. [10]	CNN	83.00%
Our Approach		$\mathbf{96.37\% \pm 0.8\%}$

Table II
COMPARISON WITH OTHER EXISTING APPROACHES (KDEF)

Methods	Classifiers	Recognition Rate
Yaddaden et al. [5]	KNN	79.69%
Rao et al. [7]	Adaboost	74.05%
Shin et al. [12]	CNN	59.15%
Our Approach		$\mathbf{90.62\% \pm 1.60\%}$

Table III
CONFUSION MATRIX IN PERCENTAGES FOR CK+ DATASET

	FE	SU	HA	DI	AN	SA	NE
FE	**96.69**	0.21	0.91	0.00	0.41	0.11	1.67
SU	0.11	**99.22**	0.03	0.07	0.02	0.01	0.54
HA	0.00	0.01	**99.89**	0.00	0.00	0.00	0.11
DI	0.02	0.09	0.40	**96.44**	0.38	0.00	2.67
AN	0.01	0.00	0.00	0.36	**94.09**	2.35	3.20
SA	0.00	0.00	0.00	0.35	3.24	**93.12**	3.29
NE	0.69	0.05	0.32	0.93	4.36	2.88	**90.76**

Table IV
CONFUSION MATRIX IN PERCENTAGES FOR KDEF DATASET

	FE	SU	HA	DI	AN	SA	NE
FE	**83.74**	4.81	1.96	1.00	3.93	3.41	1.15
SU	5.95	**92.14**	0.71	0.00	0.08	0.24	0.87
HA	0.15	0.11	**99.22**	0.19	0.07	0.15	0.11
DI	2.26	0.12	1.23	**88.73**	3.25	4.05	0.36
AN	3.25	0.04	1.23	6.59	**83.81**	2.74	2.34
SA	4.84	0.04	0.00	3.73	0.56	**88.85**	1.98
NE	0.00	0.24	0.00	0.04	1.79	3.06	**94.88**

Emotional Faces) [24] and it contains 4900 color images of 70 female and male subjects expressing seven different emotions. Each facial expression is photographed twice from 5 different angles. In our case, we have used only the frontal images as the used FCP extraction method is more efficient with this type of image.

Generally, available CNN frameworks have two operating modes: Central Processing Unit (CPU) and Graphical Processing Unit (GPU). The second one is faster than the CPU mode but can be used only if the testing platform is compatible with CUDA. In the present work, we used the GPU mode with an NVIDIA Quadro K620 as graphical card. We also checked the possibility of implementing our approach in an embedded system with limited resources. Therefore, we used the popular Raspberry Pi 2 Model B. The optimization which has been made to the CNN architectures has allowed us to use the embedded system with CPU mode for recognizing emotions from facial expressions.

B. Obtained Results

In order to evaluate the performances of our proposed approach of facial expression recognition based on CNN, we have used two different metrics. The first one consists in the *global recognition rate* and describes the accuracy of the model for each dataset by a unique accuracy value (Table I & II). The other metric is the *confusion matrix* and provides more details concerning the recognition rate for each emotion (Table III & IV).

We notice from Table I & II that our approach outperforms both of classical and DL techniques in terms of recognition rates for both datasets. Table III & IV represent the recognition rate of each emotion for both datasets. They also show that the best values are reached for SU (Surprise) and HA (Happiness) emotions. For the Cohn-Kanade Extended dataset, the recognition rate for each emotion is greater than 90.00% with an average of 95.74%. The highest value is nearly perfect and it is equal to 99.89%. Concerning the KDEF dataset, the values are lower in comparison with the first one but still considerable with an average of 90.20% and a highest value equals to 99.22%.

C. Disucssion

The efficiency of our approach cannot be validated only based on the recognition rate as shown in Table I & II, but also in terms of depth and convergence. These two criteria are specific to DL techniques. In fact, the depth is an important parameter because the more layers we have, the more complex processing is required. As a consequence, it increases the accuracy in terms of recognition. The approach proposed by Mayya et al. [11] is close to ours in terms of recognition but it uses a deeper architecture (DCNN). However, if the targeted environment is constrained in terms of hardware resources, the implantation might be difficult. The other parameter consists in convergence and the number of iterations before reaching the highest accuracy. It defines the needed time to train the CNN model. In our case, we have made some optimizations in order to reach convergence within approximately 150 iterations which is considered as time efficient. Given the results shown in Table III & IV, SU (Surprise) and HA (Happiness) emotions are easy to distinguish from the other ones due to the high accuracy when recognizing them.

Based on the obtained results using the different metrics, our proposed approach of facial expressions recognition might be adapted and exploited for further experimentations. Their main purpose is to verify the presence of a correlation between error making during ADLs and facial expressions. There might be specific new facial expressions associated to the reactions when making mistakes. The capabilities of the proposed CNN architecture in terms of depth and

convergence allow it to be used in constrained environments such as embedded systems and placed in smart homes for assistance.

VI. Conclusion

In this paper, we introduced an approach to recognize emotions from facial expressions based on a DL technique. We proposed a CNN architecture with a specific configuration in order to reach a high recognition rate in a constrained environment. Our proposed approach which combines the DL technique with image pre-processing steps has been successfully implemented in an embedded system. We are planning to integrate this approach into a complete intelligent assistance system to detect errors in daily activities using facial expressions.

References

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[2] A. Mehrabian, *Communication without words*, 2nd ed., 1968, pp. 51–52.

[3] P. Ekman, "Are there basic emotions?" *Psychol Rev*, vol. 99, no. 3, pp. 550–553, Jul. 1992.

[4] F.-S. Hsu, W.-Y. Lin, and T.-W. Tsai, "Facial expression recognition using bag of distances," *Multimedia tools and applications*, vol. 73, no. 1, pp. 309–326, 2014.

[5] Y. Yaddaden, A. Bouzouane, M. Adda, and B. Bouchard, "A new approach of facial expression recognition for ambient assisted living," in *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, 2016, p. 14.

[6] L. Zhong, Q. Liu, P. Yang, B. Liu, J. Huang, and D. N. Metaxas, "Learning active facial patches for expression analysis," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2562–2569.

[7] Q. Rao, X. Qu, Q. Mao, and Y. Zhan, "Multi-pose facial expression recognition based on surf boosting," in *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on*. IEEE, 2015, pp. 630–635.

[8] A. A. Youssif and W. A. Asker, "Automatic facial expression recognition system based on geometric and appearance features," *Computer and Information Science*, vol. 4, no. 2, p. 115, 2011.

[9] Y. Lv, Z. Feng, and C. Xu, "Facial expression recognition via deep learning," in *SMARTCOMP*. IEEE, 2014, pp. 303–308.

[10] W. Li, M. Li, Z. Su, and Z. Zhu, "A deep-learning approach to facial expression recognition with candid images," in *Machine Vision Applications (MVA), 2015 14th IAPR International Conference on*. IEEE, 2015, pp. 279–282.

[11] V. Mayya, R. M. Pai, and M. M. Pai, "Automatic facial expression recognition using dcnn," *Procedia Computer Science*, vol. 93, pp. 453–461, 2016.

[12] M. Shin, M. Kim, and D.-S. Kwon, "Baseline cnn structure analysis for facial expression recognition," in *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*. IEEE, 2016, pp. 724–729.

[13] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 675–678.

[14] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting." *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[15] B. Bouchard, S. Giroux, and A. Bouzouane, "A keyhole plan recognition model for alzheimer's patients: first results," *Applied Artificial Intelligence*, vol. 21, no. 7, pp. 623–658, 2007.

[16] C. Baum and D. F. Edwards, "Cognitive performance in senile dementia of the alzheimers type: The kitchen task assessment," *American Journal of Occupational Therapy*, vol. 47, no. 5, pp. 431–436, 1993.

[17] A. Mihailidis, J. N. Boger, T. Craig, and J. Hoey, "The coach prompting system to assist older adults with dementia through handwashing: An efficacy study," *BMC geriatrics*, vol. 8, no. 1, p. 28, 2008.

[18] C. Peters, T. Hermann, and S. Wachsmuth, "Tebra-an automatic prompting system for persons with cognitive disabilities in brushing teeth," in *Proc. of the 6th Int. Conf. on Health Informatics (HealthInf)*, 2013.

[19] E. M. Jean-Baptiste, P. Rotshtein, and M. Russell, "Cogwatch: Automatic prompting system for stroke survivors during activities of daily living," *Journal of Innovation in Digital Ecosystems*, vol. 3, no. 2, pp. 48–56, 2016.

[20] J. Broekens, M. Heerink, and H. Rosendal, "Assistive social robots in elderly care: a review," *Gerontechnology*, vol. 8, no. 2, pp. 94–103, 2009.

[21] B. N. De Carolis, S. Ferilli, G. Palestra, and V. Carofiglio, "Towards an empathic social robot for ambient assisted living." in *ESSEM@ AAMAS*, 2015, pp. 19–34.

[22] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1867–1874.

[23] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010, pp. 94–101.

[24] D. Lundqvist, A. Flykt, and A. Öhman, "The karolinska directed emotional faces (kdef)," *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet*, pp. 91–630, 1998.