Artificial Intelligence and Social Computing, Vol. 28, 2022, 39–49
https://doi.org/10.54941/ahfe1001445

Artificial Vision System to Detect the Mood of an Alzheimer's Patient

David R. Castillo Salazar[1,2], Laura Lanzarini[3], Héctor Gómez Alvarado[4], and José Varela-Aldás[5]

[1]Facultad de Ciencias Humanas de la Educación y Desarrollo Social FACHEDS, Universidad Indoamérica, Ecuador
[2]Facultad de Informática, Universidad Nacional de la Plata, Argentina
[3]Instituto de Investigación en Informática LIDI (CICPBA Center), Facultad de Informática, Universidad Nacional de la Plata, Argentina
[4]Universidad Técnica de Ambato, Ecuador
[5]SISAu Research Group, Facultad de Ingeniería y Tecnologías de la Información y la Comunicación, Universidad Tecnológica Indoamérica, Ecuador

ABSTRACT

Dementia is a brain disorder that affects older individuals in their ability to carry out their daily activities, such as in the case of neurological diseases. The main objective of this study is to automatically classify the mood of an Alzheimer's patient into one of the following categories: wandering, nervous, depressed, disoriented, bored or normal i.e. in Alzheimer's patients from videos obtained in nursing homes for the elderly in the canton of Ambato, Ecuador. We worked with a population of 39 people from both sexes who were diagnosed with Alzheimer's and whose ages ranged between 75 and 89 years of age. The methods used are pose detection, feature extraction, and pose classification. This was achieved with the usage of neural networks, the walk classifier, and the Levenshtein Distance metric. As a result, a sequence of moods is generated, which determine a relationship between the software and the human expert for the expected effect. It is concluded that artificial vision software allows us to recognize the mood states of the Alzheimer patients during pose changes over time.

Keywords: Classification, The Levenshtein distance, Pose, Neural network, Machine vision

INTRODUCTION

This research study discusses the topic of pose estimation in Alzheimer's patients, which can be defined as the mood that allows psychological categories to be identified such as nervous, depressed, disoriented, and bored, i.e. moods of elderly patients from both sexes with this neurological disease. In this case, it is Alzheimer's. Specifically for the older adult, the illness progresses at the rate of a person's age, i.e. where three individual stages have been identified: moderate, serious, and grave -- thus generating as a characteristic a change in the well-known mood disorder related to mental health, which is caused by the existence of emotional alterations in the long and short term such as sadness, or simply depression.

Research into this complex, social phenomenon in the health field raises much interest in the usage and application of new computer technologies such as artificial intelligence, as well as in terms of implementing this study in care centers for the elderly, i.e. by strengthening knowledge of the research processes with the use of computer technologies.

In the article elaborated by (Agrawal et al., 2020), an analysis has been provided together with an explanation of the usage of techniques for pose detection, i.e. in people who do exercises using yoga to help them achieve an improved posture with the use of technologies. For this purpose, images were collected from ten different yoga postures and the estimated algorithm tensorflow-pose, namely that which draws the skeleton of the human body where the angles of the joints have been extracted in real-time so as to implement various automatic learning models, e.g. by using data from 80% for training purposes and 20% for tests. This data set was tested in different machines where it learns classification models and reaches an accuracy of 99.04% by using a random forest classifier.

Author (Takano & Lee, 2020), describes a novel approach to the classification of movements of the entire body starting with human positions estimated in 2D camera images. The displacements were encoded in stochastic movement models called motion primitives. A data set was identified by registering actions performed by a subject in a laboratory with 155 movement samples. The other data set was created by registering actions carried out by older adults in care centers with 21 types of movements. The results confirmed that approximately 90% and 70% of the movements in the laboratory and the care facilities were correctly classified.

In the research carried out by (Abobakr et al., 2019), a network for the estimation and detection of deep human posture was proposed in "one single shot", i.e. as a key aspect of artificial vision, namely where SSDPose architecture is used (Single Shot Deep Pose), which is essentially a training model that detects and estimates the body position from a single image that contributes to the posture of biomechanics and ergonomics and that aids the pose estimation. The results show that posture achieves an average precision of the individual M.A.P. (Mean Average Precision) of 98.2%, an average articulation angle MAE (Mean Absolute Error) of 3.16 ± 1.23 degrees and an RMSE (Root Mean Square Error) of 4.22 ± 1.73 degrees up to 30 fps (Frames per Second).

The present text (Wint Cho et al., 2018), mentions a system that is used to improve the recognition of human actions by incorporating the Kinect sensor, i.e. in order to obtain distance characteristics of the joints. This system recognizes fundamental movements such as walking, sitting, standing, and bending. The results show better accuracy in the static K-means than in the non-static K-means.

The aforementioned works show the importance and usefulness of methods that are used to detect the mood of an Alzheimer's patient. In all of them, algorithms were used to provide solutions to different situations that arise within human activities.

METHODS AND MATERIALS

The following section presents the materials and techniques used in pose detection, extraction and classification, namely with the usage of the developed algorithms.

Methods

The methods used in this study focus on three initial aspects: First, the detection of the pose; second, the extraction of characteristics; and third, the pose classification.

Initially, a reference was made in relation to the detection of the pose of the human structure by identifying the positions in which parts of the body such as the arms, legs, head, torso, etc. are located. To do this, the Tensorflow implementation of the DeeperCut algorithm was applied, which takes an input image and detects for each individual the 17 key points. In (Munea et al., 2020), (Z. Cao, 2017), and (Li et al., 2020), information exploration methods have been presented to estimate human poses, e.g. by incorporating videos and images that utilize techniques for the collection of temporal information and that rely on the usage of new architectures for the treatment of data from one or several persons.

Feature extraction consists of revealing in an image the position of the elements that are being sought in the human body, i.e. by generating a numerical representation called feature vector. In (Sun et al., 2017), the normalization scheme is shown in two stages. The first stage is the human body and the second stage is the limbs. Also, in (G. Cao et al., 2019) a model of Multi-2D pose estimation was identified where several types of angles were extracted during the process of motion recognition.

Pose classification is the process of obtaining the spatial configtion of the parts of a body in images. The method starts with a reduction of the possible locations of the body structure using color and disparity information of the image, e.g. by using a neural network of 32 layers of inputs, 16 and 64 hidden layers, and 2 layers of outputs for classifying the pose. In (Atvar, 2018), the steps for the detection and orientation classification are performed. The detected persons are classified as belonging to the class "standing", "sitting on an object" and "sitting on the ground". A deep learning framework was used for this. The results show that the convolutional neural network produces positive results.

The Levenshtein Distance is used to classify the sequence of poses, i.e. to transform one string of characters into another. Operation is understood as either an insertion, deletion, substitution or transposition of two characters in (Thang & Huy, 2010) This process is applied in many fields such as language processing, speech recognition, information theft detection, biological computation, etc. and in (Zhao & Sahni, 2020) two strings have been used to find a sequence of edit operations of length equal to the distance with the usage of different operations.

Materials

Initially, the research process was focused on taking video information from an estimated time frame of 1 minute for 45 Alzheimer's patients, i.e. where one took as a reference men whose ages ranged from 75 to 86 and women whose ages ranged from 75 to 89 years. These data were used as an initial sample for data recording. During the validation process, 39 samples were determined as being valid for the pose detection because the video does not show the whole body or for the fact that the samples were obtained for people whose clothes cover the key points of the human structure.

In addition, a series of Google images of 221 people sitting and 222 people standing with a total of 443 images was used.

The procedure that was carried out required specific equipment consisting of a laptop with vast CPU and GPU capacity for data processing under the Linux Ubuntu operating system. With regard to the software for the development of the algorithms, we used Python 3.5 and TensorFlow 1.22 to run DeeperCut and to implement the neural network of the pose classifier.

After completing the processing stage, the results were saved in an Excel file, i.e. where data such as the "Start" time (which is the time when the patient's analysis commenced), "Processing time" (which records the total time), "Time" (which shows the processing in seconds), "Pose" (which is the statistical mode of all the results delivered by the pose classifier during a one second time period), and "Status" (which shows the mood of the patient.

The following section describes the development of the model in relation to the research.

DEVELOPMENT OF THE MODEL

The model shows the development of software that is capable of automatically classifying the mood of an Alzheimer's patient into one of the following categories: wandering, nervous, depressed, disoriented, or bored.

To achieve this, the software detects at each "t" time instant a set of 17 P_t key points on the patient's body and derives from them an X_t feature vector of 48 components. Subsequently, this vector is used to feed a classifier that assigns one of the following poses to the patient: standing, sitting, or walking. Finally, the different poses of the patient are accumulated over a T_c time in order to form a $\widehat{Q}t$ sequence, which, when compared with a set of standard Q sequences and by means of the Levenshtein distance, allows us to deduce the patient's mood.

The complete system architecture is shown in Figure 1.

Pose Detection

To detect the pose, the TensorFlow implementation of the Deep Cut algorithm was used (Insafutdinov et al., 2016) This algorithm receives the image of a patient as input and delivers as output the location of 17 key points distributed throughout the body. These points are nose, eyel, eyer, earl, earr, shoulderl, shoulderr, elbowl, elbowr, handl, handr, hipl, hipr, kneel, kneer, footl, and footr. This set is represented in abbreviated form, i.e. as $P_t = \{p_1, p_2, \ldots, p_i, \ldots, p_{17}\}$, namely where each element represents the coordinate $p_i = (x, y)$ of one of the above key points, in its respective order. This research has a unit test that allows us to detect these 17 key points for an individual.

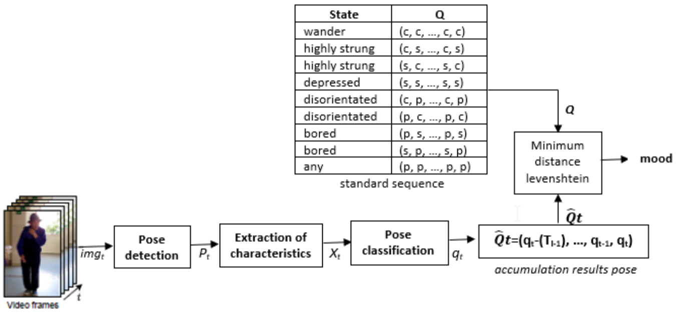

Figure 1: General arhitecture of the system.

Feature Extraction

In order to classify the pose state of a patient, it is necessary to construct a feature vector starting from the 17 key points detected by the DeeperCut algorithm in (Insafutdinov et al., 2016) and (Xu et al., 2018) is used to detect the parts of the body. For this process, we first normalize the key points with respect to the center of gravity and the body size following the methodology of [19]. In this way, the feature vector will be independent of the patient's size or distance from the camera. The head length is defined as shown in equation 1.

$$\begin{aligned} length_{\text{head}} = m\acute{a}x(&(neck - \text{eye}_l)\ (neck - \text{eye}_r)\,, \\ &(neck - \text{ear}_l)\ ,\ (neck - \text{ear}_r)\,, \\ &(nose - \text{eye}_l)\ (nose - \text{eye}_r)\,, \\ &(nose - \text{ear}_l)\ ,\ (nose - \text{ear}_r)) \end{aligned} \quad (1)$$

Pose Clasifier

The function of the pose classifier is to infer the pose of the patient at each time instant. The pose can be: sitting, standing, or walking. In (Guerra et al., 2020), (Akhil et al., 2017), and (Saho et al., 2020) similar aspects were identified. For this purpose, the pose classifier first classifies the feature vector with the help of a neural network to find out whether the patient is sitting or not. In case the patient is not seated, a second algorithmic stage analyzes the patient's gait to know whether the patient is standing or walking.

Neural Network

The first stage of the pose classifier is a multilayer neural network that takes as input the feature vector $X_t \in \mathbb{R}^{48}$ and delivers as output a two-dimensional vector$\widehat{y}_t = [\widehat{y}_1, \widehat{y}_2]$, i.e. where $\widehat{y}_1$ y $\widehat{y}_2$ correspond to the probability of sitting or non-sitting, respectively. The schematic of this neural network is shown in (Figure 2). In (Hoa & Bui, 2016) a study of human body posture classification using a fuzzy neural network of two neurons was carried out.

For the training of this network, two sets of images were collected from Google Images: the first set, which is called sitting, consists of 221 images in

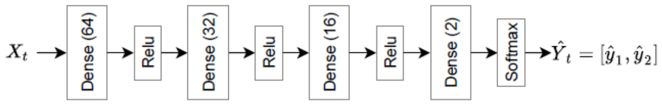

Figure 2: The architecture of the neural network is responsible for classifying the patient's pose in the classes: sitting or not-sitting. The number of neurons in the first, second, third, and fourth layers is 64, 32, 16 and 2, respectively. The activation function used in the input layer and in the hidden layers is the Relu function and the one used in the output layer is the Softmax function.

which only standing people appear. The second set, which is called standing, consists of 222 images in which only standing people appear.

In (Insafutdinov et al., 2016) and (Xu et al., 2018) were then processed with the DeeperCut algorithm in order to extract the features of people from each set of images. The set of key points obtained was filtered in order to eliminate those where it was not possible to detect one of the key points from the head, i.e. at least two from the torso, or at least 3 from the legs. In this way, 663 feature vectors of seated persons and 656 feature vectors of standing persons were obtained. From among these vectors, 80% were randomly chosen for training and the remaining 20% for validation. In the training, the neural network was trained to minimize the following error function.

Walking Classifier

The function of the walk classifier is to infer whether a patient is standing or walking, when previously it has been ruled out that the patient was sitting. To develop this algorithm, we researched the variation over time of various angles between different parts of the body, i.e. in order to find one that would enable us to accomplish this task. Although this task may appear to be simple, e.g. analyzing the angle formed between a patient's legs, it is not, due to the following factors:

- Elderly people often walk slowly, so detecting the variation of an angle due to their movement can be difficult.
- Oftentimes the angle variations, which are due to variations in the key point detection by the DeeperCut algorithm, are much larger than the patient's walks. This is due to several factors, but among them, we can highlight the variations that cause the detected coordinates to have small random changes around a central point.
- In this particular research study, the videos were captured by a moving camera; this introduces a variation that should not be taken into account.

Levenshtein Distance

In order to know the mood of the patient, the different results of the pose classifier were accumulated over time. By default, it is processed at 10 fps. Thus, for every second you obtain 10 pose results. As many of these values might be repeated or there might be some errors, the mode is extracted from each second to obtain a single result. Then, by accumulating the results of each second, a $\widehat{Q}_t$ sequence is constructed for each t with the results of the last T_c seconds. This sequence is compared with each of the standard T_c. A variation of the Levenshtein Distance, which is known as Damerau - Levenshtein (Boytsov, 2011), (Behara et al., 2020) and (Gabrys et al., 2018) - is used for this comparison. This distance computes the number of operations required to transform a sequence *a* into another sequence b by means of the following four operations: delete, insert, replace, and transpose. Mathematically, the Damerau - Levenshtein distance between two sequences *a* and *b*.

The following section displays the results of the research that was carried out.

RESULTS

To evaluate the performance of the software, 39 test videos were processed. Next, the sequence of mood states output by the software was compared with the sequence of mood states annotated by a medical expert. Prior to testing, the software was conFigd as follows.

- The processing rate was set to fps = 10, i.e. for every second we obtained 10 results for a patient's pose. The final result for each second is the statistical mode of these 10 results.
- The length of the standard sequences was set at T_c=5, i.e. each standard sequence is of length 5. So, for example, the standard sequence for "wandering" is is ccccc, whereas for "nervous" it is scscs. The mood result delivered by the software for each second is the name of the standard sequence with the smallest Damerau - Levenshtein distance in relation to the sequence formed by the concatenation of the last 5 results (of the last 5 seconds).
- All videos were resized to 320x240 pixels.

By using the above conFigtion and given that there was a length T_test video (in seconds), it is processed to produce a sequence of moods $\widehat{S} = \{\widehat{Q}_1, \ldots, \widehat{Q}_t, \ldots, \widehat{Q}_T\}$, i.e. where $\widehat{Q}_t$ is the mood at second t . This sequence is then compared with the true sequence $S = \{Q_1, \ldots, Q_t, \ldots, Q_T\}$ which is noted by a medical expert. The comparison process is as follows: at every t second if $\widehat{Q}_t$ and Q_t are equal, there is a true positive (VP), or, conversely, there is a false negative (FN). With the results obtained after processing all the videos, confusion matrix was constructed. Confusion matrix constructed with the 39 test videos. Here E1, E2, E3, E4, E5 correspond to the classes of wandering, nervous, depressed, disoriented and bored, respectively.

In order to obtain a global metric of the software performance, the Macro - average - accuracy - score (MacroAvg) (Xia et al., 2018) (Yang et al., 2011) (Zhang et al., 2018), is calculated, which is defined as the arithmetic mean of all the accuracy scores of the different classes, namely equation 2.

$$MacroAvg = \frac{PrecE_1 + PrecE_2 + PrecE_3 + PrecE_4 + PrecE_5}{5} \tag{2}$$

Accuracy for the wandering class ($E_1 = 0.86$); Accuracy for the nervous class ($E_2 = 0.25$); Accuracy for the depressed class ($E_3 = 0.73$); Accuracy for the wandering class ($E_4 = 0.20$); Accuracy for the bored class ($E_5 = 0$). By substituting the above results into equation 2, the following MacroAvg was obtained: MacroAvg=0.43

The accuracy for the bored and nervous classes is low, which subsequently affects the MacroAvg. However, it should be noted that there are very few examples for these classes in the test videos. Consequently, their results are not representative. If we excluded these classes from the confusion matrix, we would obtain a MacroAvg equal to:

$$\frac{0.86 + 0.73 + 0.2}{3} = 0.6 \tag{3}$$

The boring and nervous classes are difficult to separate. Note in the confusion matrix that 5 results that should have been classified as bored were classified as nervous. This is because the pose of many patients is misinterpreted as walking when they are actually standing. Therefore, the sequence pattern of a bored *pspsp* patient ends up becoming *cscsc,* which has a smaller distance to the sequence pattern of the nervous class than to the sequence pattern of the bored class. A similar problem arises with the classes of wandering and disoriented, namely where the sequence pattern (Dolatabadi et al., 2017) *pcpcp* ends up becoming *ccccc*. The confusion that arises when identifying someone standing as walking is due to the fact that the videos were captured with a moving camera, which induces an error in the calculation of the angle between the vector v and the vector u.

Several videos identified problems to properly classify the pose due to the fact that the patients were dressed in suits and clothing accessories such as scarves or ruanas that make it difficult to detect key points.

The software recognizes a patient's mood every second based on the Damerau - Levenshtein distance, so it does not require training with previously labeled mood sequences. For this reason, it can be said that the software works in an unsupervised mode.

The following section presents the conclusions of the research that was carried out.

CONCLUSION

This paper presents the development of computer vision software that is capable of recognizing the mood of an Alzheimer's patient such as wandering, nervous, depressed, disoriented or bored, which is based on the changes of their pose over time. To characterize the pose, we first detected a series of 17 key points on the patient's body (by using the DeeperCut algorithm) and then calculated some angles and lengths of their joints and body parts, respectively. Following this procedure, a 48-component feature vector was obtained, which feeds a neural network that is responsible for classifying the patient's pose such as sitting or non-sitting.

When a patient is classified as non-sitting, there is an additional procedure that classifies the patient's pose such as standing or walking. Finally, the different poses of the patient are accumulated over a Tc time frame, i.e. in order to form a $\widehat{Q}_t$ sequence, which, when compared to a set of standard Q, sequences by means of the Damerau - Levenshtein distance, enables us to deduce the mood of the patient.

The software is evaluated with 39 videos in which the patient's mood was labeled by a medical expert. For this test, the software was conFigd as follows: the processing rate was set to fps $= 10$, the length of the standard sequences was set to $T_c = 5$ and all the videos were resized to 320×240 pixels. The accuracy obtained for the wandering class was 0.86, for the nervous class 0.25, for the depressed class 0.73, for the disoriented class 0.2, and for the bored class 0. The calculation of the arithmetic average of all accuracies yielded a macro-average-precision-score of 0.43.

REFERENCES

Abobakr, A., Abdelkader, H., Iskander, J., Nahavandi, D., Saleh, K., Attia, M., Hossny, M., & Nahavandi, S. (2019). SSDPose: A single shot deep pose estimation and analysis. *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, *2019-Octob*, 1862–1868. https://doi.org/10.1109/SMC.2019.8913964

Agrawal, Y., Shah, Y., & Sharma, A. (2020). Implementation of machine learning technique for identification of yoga poses. *Proceedings - 2020 IEEE 9th International Conference on Communication Systems and Network Technologies, CSNT 2020*, 40–43. https://doi.org/10.1109/CSNT48778.2020.9115758

Akhil, V. M., Varghese, J., Rajendrakumar, P. K., & Sivanandan, K. S. (2017). Artificial neural network based study of torque at knee during sit to stand and back to sit movements. *2016 International Conference on Control Instrumentation Communication and Computational Technologies, ICCICCT 2016*, 720–724. https://doi.org/10.1109/ICCICCT.2016.7988046

Atvar, A. (2018). *Insan Duru Â¸ s ve Yönelimlerinin Derin Ö Ë˜ grenme ile Sınıflandırılması Classification of Human Poses and Orientations with Deep Learning*. 2–5.

Behara, K. N. S., Bhaskar, A., & Chung, E. (2020). A novel approach for the structural comparison of origin-destination matrices: Levenshtein distance. *Transportation Research Part C: Emerging Technologies*, *111*(May 2019), 513–530. https://doi.org/10.1016/j.trc.2020.01.005

Boytsov, L. (2011). Indexing methods for approximate dictionary searching: Comparative analysis. *ACM Journal of Experimental Algorithmics*, *16*(1). https://doi.org/10.1145/1963190.1963191

Cao, G., Pu, Y., & Li, Y. (2019). Human Motion Capture Using a Multi-2D Pose Estimation Model. *2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, *1*, 64–67. https://doi.org/10.1109/IHMSC.2019.00023

Cao, Z. (2017). *Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields âˆ—*. 1302–1310. https://doi.org/10.1109/CVPR.2017.143

Dolatabadi, E., Taati, B., & Mihailidis, A. (2017). An automated classification of pathological gait using unobtrusive sensing technology. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *25*(12), 2336–2346. https://doi.org/10.1109/TNSRE.2017.2736939

Gabrys, R., Yaakobi, E., & Milenkovic, O. (2018). Codes in the Damerau Distance for Deletion and Adjacent Transposition Correction. *IEEE Transactions on Information Theory*, *64*(4), 2550–2570. https://doi.org/10.1109/TIT.2017.2778143

Guerra, B. M. V., Ramat, S., Gandolfi, R., Beltrami, G., & Schmid, M. (2020). Skeleton data pre-processing for human pose recognition using Neural Network. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2020-July*, 4265–4268. https://doi.org/10.1109/EMBC44109.2020.9175588

Hoa, N. T., & Bui, T. D. (2016). Classifying human body postures by a two-neuron fuzzy neural network. *2016 IEEE RIVF International Conference on Computing and Communication Technologies: Research, Innovation, and Vision for the Future, RIVF 2016 - Proceedings*, 142–146. https://doi.org/10.1109/RIVF.2016.7800284

Insafutdinov, E., Pishchulin, L., Andres, B., Andriluka, M., & Schiele, B. (2016). Deepercut: A deeper, stronger, and faster multi-person pose estimation model. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9910 LNCS*, 34–50. https://doi.org/10.1007/978-3-319-46466-4_3

Li, Y., Li, K., Wang, X., Yi, R., & Xu, D. (2020). *Exploring temporal consistency for human pose estimation in videos*. *103*. https://doi.org/10.1016/j.patcog.2020.107258

Munea, T. L., Jembre, Y. Z., & Weldegebriel, H. T. (2020). *The Progress of Human Pose Estimation : A Survey and Taxonomy of Models Applied in 2D Human Pose Estimation*. *XX*. https://doi.org/10.1109/ACCESS.2020.3010248

Saho, K., Inuzuka, K., & Shioiri, K. (2020). Person Identification Based on Micro-Doppler Signatures of Sit-to-Stand and Stand-to-Sit Movements Using a Convolutional Neural Network. *IEEE Sensors Letters*, *4*(3), 7–10. https://doi.org/10.1109/LSENS.2020.2975219

Sun, K., Lan, C., Xing, J., Zeng, W., Liu, D., & Wang, J. (2017). Human Pose Estimation Using Global and Local Normalization. *Proceedings of the IEEE International Conference on Computer Vision*, *2017-Octob*(d), 5600–5608. https://doi.org/10.1109/ICCV.2017.597

Takano, W., & Lee, H. (2020). Action Description from 2D Human Postures in Care Facilities. *IEEE Robotics and Automation Letters*, *5*(2), 774–781. https://doi.org/10.1109/LRA.2020.2965394

Thang, D. Q., & Huy, P. T. (2010). Determining restricted Damerau-Levenshtein edit-distance of two languages by extended automata. *2010 IEEE-RIVF International Conference on Computing and Communication Technologies: Research, Innovation and Vision for the Future, RIVF 2010*. https://doi.org/10.1109/RIVF.2010.5632914

Wint Cho, T. Z., Win, M. T., & Win, A. (2018). Human action recognition system based on skeleton data. *Proceedings - 2018 IEEE International Conference on Agents, ICA 2018*, 93–98. https://doi.org/10.1109/AGENTS.2018.8458495

Xia, X., Liu, J., Yang, T., Jiang, D., Han, W., & Sahli, H. (2018). Video Emotion Recognition using Hand-Crafted and Deep Learning Features. *2018 1st Asian Conference on Affective Computing and Intelligent Interaction, ACII Asia 2018*, 1–6. https://doi.org/10.1109/ACIIAsia.2018.8470326

Xu, H., Shen, L., Zhang, Q., & Cao, G. (2018). Fall behavior recognition based on deep learning and image processing. *International Journal of Mobile Computing and Multimedia Communications*, *9*(4), 1–15. https://doi.org/10.4018/IJMCMC.2018100101

Yang, W., Feng, Q., Lu, Z., & Chen, W. (2011). Metric learning for maximizing MAP and its application to content-based medical image retrieval. *Proceedings - International Symposium on Biomedical Imaging*, 1901–1904. https://doi.org/10.1109/ISBI.2011.5872780

Zhang, X., Xu, M., & Zheng, T. F. (2018). Ensemble System for Multimodal Emotion Recognition Challenge (MEC 2017). *2018 1st Asian Conference on Affective Computing and Intelligent Interaction, ACII Asia 2018*, *Mec 2017*, 1–6. https://doi.org/10.1109/ACIIAsia.2018.8470352

Zhao, C., & Sahni, S. (2020). Linear space string correction algorithm using the Damerau-Levenshtein distance. *BMC Bioinformatics*, *21*, 1. https://doi.org/10.1186/s12859-019-3184-8