First-person camera system to evaluate *Tender Dementia-care* skill

Atsushi Nakazawa
Graduate School of Informatics,
Kyoto University
nakazawa.atsushi@i.kyoto-u.ac.jp

Miwako Honda
Geriatric Research Division,
Tokyo Medical Center
honda-1@umin.ac.jp

Abstract

In this paper, we describe a wearable first-person video (FPV) analysis system for evaluating the skill levels of tender dementia-care technique. Using this system, caregivers can evaluate and elevate their care levels by themselves using the systems' feedbacks. From the FPVs of care sessions taken by wearable cameras worn by caregivers, we obtained the 3D facial distance, pose and eye-contact states between caregivers and receivers by using facial landmark detection and deep neural network (DNN)-based eye contact detection. We applied statistical analysis to these features and developed algorithms that provide scores for tender-care skill. To find and confirm our idea, we conducted chronological study to observe the progression of tender care-skill learning using care learners. First, we took FPVs while care training scenes involving novice caregivers, tender-care experts and middle-level students, and found major behavioural differences among them. Second, we performed the same experiments for the participants before and after training sessions of the care. As the result, we found the same behavioural difference between 1) novices and experts and 2) novices before and after taking training sessions. These results indicate that our FPV-based behavior analysis can evaluate the skill progression of the tender dementia-care.

1. Introduction

As the elderly population increases, the number of people suffering from dementia continues to grow. As a result, the care that needs to be administered to them is becoming increasingly important in social terms [32, 20, 7].

Dementia occurs when the brain is damaged by maladies such as Alzheimer's disease and Lewy body dementia and produces a set of symptoms that include memory loss and difficulties with thinking, problem-solving, and verbal communication. Dementia can be accompanied by psychosis, agitation and aggression; thus, caring for people with dementia is quite difficult [6].

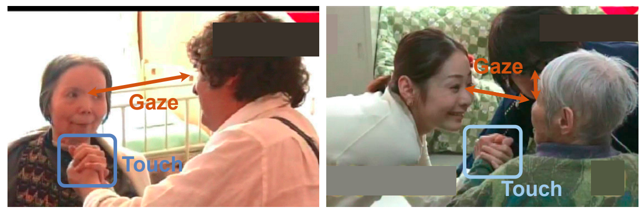

Figure 1. Example pictures of the tender-care skill *Humanitude* technique. *Humanitude* is a multi-modal care-communication method that combines gaze (keeping eye-contacts up close), touching and talking skills, then realizes smoother communication with people with dementia (PwD). As the result, introducing the technique reduces the events of BPSD (Behavioral and Psychological Symptoms of Dementia) and caregiver's burden.

Problems in dementia-care: The burden of caregiver for dementia patients have been reported in a lot of literature [19, 9, 2, 31, 10, 13, 12, 26]. According to the recent meta-review paper [2], the larger caregivers burden is related to 1) female sex, 2) low education, 3) cohabitation with care recipient, 4) caregiving time and effort, 5) financial stress, 6) lack of choice and inability to continue regular employment. As the result, caregiver tends to having larger risk in mortality, weight loss, poor self-care and sleep deprivation.

Effects of interventions for reducing caregiver's burden have been reported as well [25, 16, 1, 24, 11, 12]. Interventions are categorized into several types: psychoeducational intervention, psychosocial intervention, cognitive behavior therapy, respite, caregiver support groups, anticholinergic and antipsychotic drugs, and skill training. As the results, practical interventions to reduce caregiver's burden are 1) encouraging caregivers to function as a member of the care team, 2) encouraging caregivers to improve self-care and maintain their health, 3) providing education and information, 4) coordinating for assistance with care, 5) encouraging caregiver to access respite care and 6) using the supports of technology [2]. Specifically, there are several reports that skills training such as coping skill training (CST) may reduce the caregiver strain, depression and fatigue in caregivers of the patient with cancers [24, 8].

***Humanitude* tender-care style**: As one of the intervention approach to reduce the caregiver's burden, we are studying about the caregiving style *Humanitude*, which has been spotlighted by care professionals and family caregivers since it can reduce the occurrence of BPSD (Behavioral and Psychological Symptoms of Dementia) events and caregivers' burden [17]. Humanitude was developed by Y. Geneste and R. Marescotti 35 years ago [15] and has been introduced in more than 600 hospitals and nursing homes in Europe. Humanitude primarily uses a combination of four communication skills: gaze, verbal communication, touch, and helping care receivers to stand up. Example pictures in Humanitude care are shown in Figure 1. Several studies have reported that the cost-efficiency of introducing Humanitude is around 20 times that of care without it because of a 40

First person camera system to evaluate tender-care skill: In this paper, we show our studies to understand the tender-care skill through the behavior quantization using first person video (FPV) analysis. The system obtain the face-to-face postures and eye contact states between a caregiver and a receiver from FPV, and then infers the skill level from the feature vectors of FPV. Using the system's outputs such as individual skill levels and difference to the experts, learners can perform self-training and raise their tender-care skills. The contributions of this paper are:

1. We show the FPV can differentiate and thus evaluate the skill-level of tender-cares, which depends on the mutual facial distances and poses between a caregiver and a receiver.
2. We show a prototype system that can be used for novice care learners who want to raise their care skills by themselves.
3. In contrast to our previous work [23], we newly introduced reliable face detectors and performed a chronological study that shows our system can visualize not only the skill levels, but also skill progression of the learners.

2. Related work

In this section, we show related studies regarding care skill evaluation and first-person video-based skill evaluations.

Care skill quantization: There are several approaches that use care skill quantization. In computer science, Ishikawa et al. developed a method of care skill evaluation based on the knowledge of care experts [18]. They categorized care skills into three layers: intramodality, intermodality and multimodal-interaction. Intramodality consists of behavior primitives such as gaze, speech, touch, nodding and knocking on a door. Intermodality shows the relationships among intramodalities, such as comprehensiveness of care, waiting for elderly people's actions and consistency. Multimodal-interaction consists of actions that develop a relationship between actors, such as eye contact and verbal/nonverbal dialogue. They also developed a web interface that shows care learners care skills in visual form to confirm the effectiveness of the system.

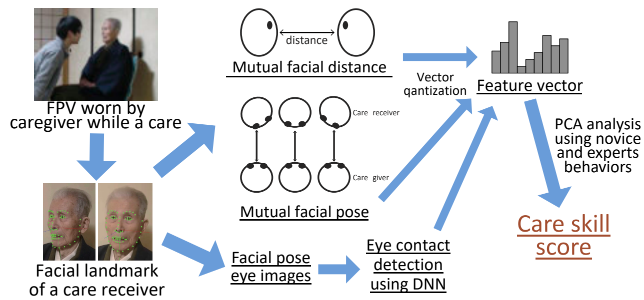

Figure 2. Overview of the care skill estimation method using first-person video.

First-person video analysis for skill evaluation: There have been a number of studies on action recognition and prediction using FPVs [14, 30, 28, 29, 22]. However, few studies have been conducted for *skill* evaluation. In recent years, Bertasius et al. showed a method to assess a basketball player's performance from FPVs. They designed and used temporal CNN and long short-term memory (LSTM) architecture to evaluate whether a particular play in basketball was good or not from a player's FPV [5]. In the medical field, Hei et al. proposed a method for evaluating skill in robotic surgical operations from video images. Their method tracks the keypoints of surgical robot instruments by using cloud sourcing or hourglass networks and evaluates the skill by support vector machine analysis [21].

3. Proposed method

The flow of our skill evaluation is illustrated in Figure 2. From a first-person camera worn by a caregiver we obtained mutual facial distances, mutual facial poses and eye contact states. Then, we estimated tender-skill scores through an unsupervised analysis. In the following subsections, we first describe the first-person camera hardware and then our algorithms we used for analyzing FPVs.

3.1. Hardware

We used a Pupil Labs camera system [27] as the head-mounted first-person camera. Since we assume the application in real care training condition, we did not used the eye-gaze-tracking functionality of the system since the calibration is troublesome and the eye camera is sometimes unsafe for caregiver's eyes. The Pupil Labs system have two types of frontal camera: one is narrow view (about 60 deg.) and another is spherical view. Regarding the videos taken by latter types of camera, we convert them to the view of

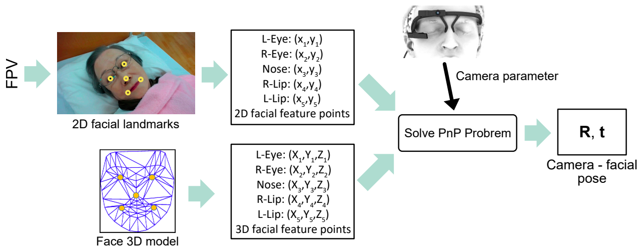

Figure 3. 3D pose estimation using the response of Amazon Rekognition (Facial detection from Video). Since Amazon Rekognition only returns the 2D coordinates of five facial landmark points, we obtain 3D facial positions by solving PnP problem using a 3D facial model and the camera parameter of FPV camera.

the former camera by using the image mapping techniques with their camera parameters.

3.2. Face detection and 3D pose estimation

We then obtained facial positions, poses and eye locations from the input FPVs. In here, we used the OpenFace library [4] or Amazkon Rekognition (Face recognition) [3]. For the case using OpenFace, facial 3D position and facial 68 landmark points are directly obtained. For the case using Amazon Rekognition, we obtain 3D facial positions from five facial landmark points given from Amazon Rekognition (Fig. 3). Namely, using five 2D facial landmark points, corresponding 3D landmark points from a standard 3D facial model and the camera parameter of FPV camera, we solve the Perspective-n-Point (PnP) problem and obtain the 3D rotation and translation between the face and the camera. These information are directly used for the skill analysis and the further eye contact detection step.

3.3. Histograms of facial distances and poses

To quantize the face-to-face communication behaviors between caregivers and care receivers, we encoded the mutual facial distances and poses obtained from OpenFace as illustrated in Figure 4. We computed the histograms $\mathbf{f}_{\text{dist}} = [f^1_{\text{dist}}, \ldots, f^{11}_{\text{dist}}]$, $\mathbf{f}_{r_x} = [f^1_{r_x}, \ldots, f^9_{r_x}]$, $\mathbf{f}_{r_y} = [f^1_{r_y}, \ldots, f^9_{r_y}]$ and $\mathbf{f}_{r_z} = [f^1_{r_z}, \ldots, f^9_{r_z}]$ that represent the mutual facial distances and poses from all frames in a care session. The bins were set to every 0.1 [m] from 0.0 to 1.0 [m] for the distance feature and 20 [deg] from -90 to +90 [deg] for angular features. The distances larger than 1.0 [m] were voted to the last bin. Thus, for example, f^1_{dist} indicates the number of frames where the mutual facial distances were from 0.0 to 0.1 [m] and $f^4_{r_x}$ indicates the number of frames where the mutual facial rotation (r_x) was from -30 to -10 [deg].

3.4. Visualization

After obtaining histograms, we normalized the histogram and applied principal component analysis (PCA). While many data analysis or machine learning techniques

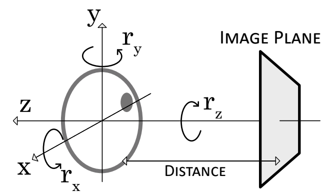

Figure 4. Facial pose parameters.

have been proposed, we used PCA for its simplicity and reliability in exploratory data analysis (EDA). Since we tried to find tender-care technique skills in a bottom-up (data-driven) manner, this nature of PCA fitted our task better than other more complicated methods such as non-linear or supervised-learning based approaches.

We here denote $\mathbf{f}^{\mathrm{s}}$ as a $\mathrm{D} \times 1$ column vector that represents a normalized histogram of either $\mathbf{f}_{\text{dist}}$, $\mathbf{f}_{r_x}$, $\mathbf{f}_{r_y}$ or $\mathbf{f}_{r_z}$ of a subject $\mathrm{s} \in \{1, \ldots, \mathrm{M}\}$. Histograms of all subjects can be decomposed by using $\mathrm{D} \times 1$ column eigenvectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_{\mathrm{M}}\}$ and eigenvalues $\{\lambda_{1,1}, \ldots, \lambda_{\mathrm{D,M}}\}$, where D is the dimension of the histogram:

$$\begin{bmatrix} \mathbf{f}^1 & \cdots & \mathbf{f}^{\mathrm{M}} \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_{\mathrm{D}} \end{bmatrix} \begin{bmatrix} \lambda_{1,1} & \cdots & \lambda_{1,\mathrm{M}} \\ \vdots & \ddots & \vdots \\ \lambda_{\mathrm{D},1} & \cdots & \lambda_{\mathrm{D,M}} \end{bmatrix}.$$

We plot the eigenvalues of all subjects to visualize the distribution of their behaviors, as well as to analyze the elements of eigenvectors to find the relation between skill levels and behavioral features.

3.5. Eye contact features

Another feature is counting eye contact bids, which was introduced by Ye et al. [33]. They assume eye contact bids, i.e., situations when a subject wearing an FPV camera is gazed at by other subjects. Since the definition of eye contact is making mutual eye gazetwo people look at each other at the same time – eye contact bids are not the same as actual eye contact. If we want to accurately detect eye contact, two people must wear FPV cameras or a caregiver must use an eye gaze tracker (EGT) device that detects observers gaze information. However, from the practical point of view it is difficult to use two FPV cameras or an EGT device since 1) it is difficult for subjects with dementia to wear such devices and 2) even for caregivers it is difficult to use eye trackers in actual care scenes due to their noticeable appearance, calibration requirements and headmount drift. Therefore, rather than accurately detecting eye contact, we tried to measure and use eye contact bids for evaluating care skills. We used facial poses and eye images for detecting eye contact bids using DNNs developed by our group that utilize facial position, eye-region images and their temporal changes [23].

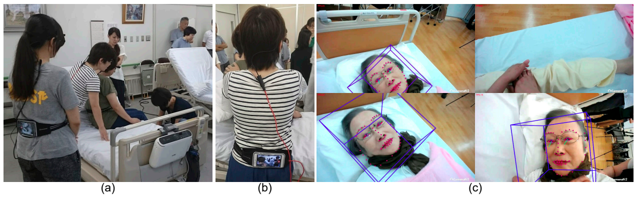

Figure 5. Experimental scenes and example FPV frames of care learning scenes. (a) Data capturing, (b) first-person camera and recorder (Pupil Labs + smartphone) and (c) example frames (Facial annotations overlaid).

4. Dataset

The took FPV dataset obtained in care learning scenes. We recorded the FPV videos equipped to the caregiver during Humanitude care teaching classes.

4.1. FPVs in care scenes with different skill levels

To find the behavioral difference between novices and experts, we prepared first-person videos of a) two Humanitude care experts (instructors), b) seven middle-level Humanitude caregivers and c) three novice Humanitude caregivers as shown in Table 1 and Figure 5. In all videos, caregivers were equipped with the Pupil Labs first-person camera and performed the same task (Figure 6):

Step 1 Approach the simulated patient while making eye contact,

Step 2 Perform the care, namely, move turn over the patient's body and clean the back, and

Step 3 Leave the care receiver.

4.2. FPV in care scenes with chronological difference

Since we want to observe novices' skill progressions while learning Humanitude technique, we took FPVs of the care novices of a) before taking Humanitude care learning sessions and b) one-week later. We asked 12 participants to perform the same task as the dataset of 4.1 and obtain their FPVs.

5. Experiments

The first experiment was performed for an actual Humanitude care training scene. In it, we obtained data from a novice caregiver and a Humanitude care expert and compared the results through the use of an unsupervised learning algorithm. In the second experiment, we obtained the data of 12 novices before and after one-week training session, and compared the differences.

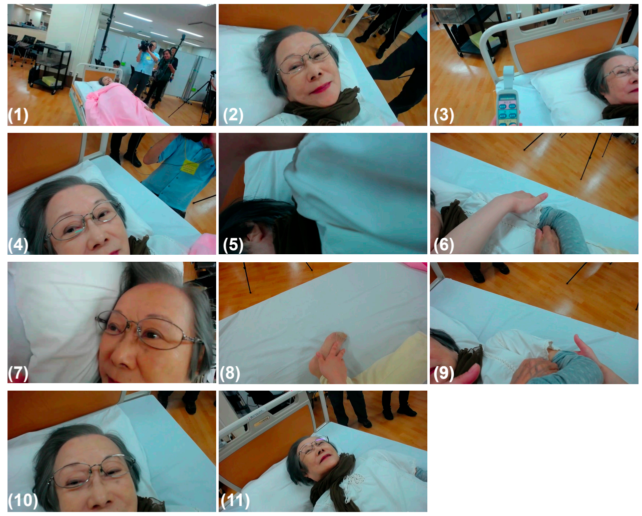

Figure 6. An example flow of the Humanitude care used int the dataset. (1)-(4): Approach the simulated patient while making eye contact, (5)-(9): perform the care (turn over the body), and (10)-(11): greet and leave the care receiver.

5.1. Experiment 1: Difference of face-to-face communication between experts, middle levels and novices

In the first experiment we compared the occurrence of Humanitude care skill between novice and expert caregivers using the FPVs of care learning scenes. We obtained the number of eye contact frames, mutual facial distances and poses from a care scene dataset and compared the results.

Analysis and results: The occurrences of eye contact frames, average mutual facial distance and poses (angles) are shown in Table 1 and normalized histograms of each feature are shown in Figure 7 (left). As seen from the table, the facial detection rate of the amazon rekognition is higher than that of OpenFace for all videos.

For analyzing the facial mutual distance and pose, we applied PCA to the histograms. The resulting PCA scores are shown in Figure 7 (right), where the x-axis shows the scores of the first component and the y-axis indicates the scores of the second PCA component. From the eye contact rates and PCA analysis results, we were able to clearly distinguish the scores of novices and experts for eye contact rate, mutual facial distance and r_z PCA scores. There were significant differences in eye contact rate between the expert & middle-level and novice groups ($p = 0.0452$), and clear thresholds at about $x = 0.16$ for mutual facial distance and at about $x = -0.18$ for the r_z PCA scores. In the mutual facial distance category, the histograms showed that the expert caregivers and most of the middle-level ones approached the care receiver such that the distance was less than 30 [cm]. In the mutual facial pose category, there were clear dissimilarities in the z-rotation, which is the rotation

Table 1. The result of Experiment 1.

Name	Total frames	Face detected (OpenFace)	Face detected (Amazon Rekog.)	Eye contacted frames	Av. mutual facial distance [mm]	Av. mutual facial pose [deg]		
						r_x	r_y	r_z
Expert A	3442	2148 (62.4%)	2887 (83.9%)	2637 (76.6%)	388.76	-0.90	-20.29	-23.75
Expert A	4851	2070 (42.7%)	3468 (71.5%)	3032 (62.5%)	668.60	1.940	-3.90	-18.62
Middle A	3214	1857 (57.8%)	2324 (72.3%)	1587 (49.4%)	533.89	3.49	-13.03	-28.77
Middle B	2659	1248 (46.9%)	1660 (62.4%)	1511 (56.8%)	475.06	5.35	-11.63	-40.13
Middle C	2024	1171 (57.9%)	1765 (87.2%)	1254 (62.0%)	461.81	-1.06	-9.12	-29.61
Middle D	2775	1108 (39.9%)	1875 (67.6%)	1153 (41.5%)	580.13	4.44	-13.84	-30.62
Middle E	4506	2547 (56.5%)	3530 (78.3%)	2421 (53.7%)	616.63	6.80	-13.36	-38.09
Middle F	3062	2176 (71.1%)	2859 (93.4%)	2041 (66.7%)	334.40	-5.79	-17.98	-22.33
Middle G	2485	683 (27.5%)	1646 (66.2%)	1126 (45.3%)	680.19	-0.57	-4.55	-29.79
Novice A	6287	1648 (26.2%)	3642 (57.9%)	842 (13.4%)	902.94	-0.72	10.52	-43.14
Novice B	6168	2675 (43.4%)	3642 (59.0%)	941 (15.3%)	902.94	-0.72	10.52	-43.14
Novice C	4710	1102 (23.4%)	2180 (46.3%)	764 (16.2%)	940.46	4.07	-9.83	-26.27

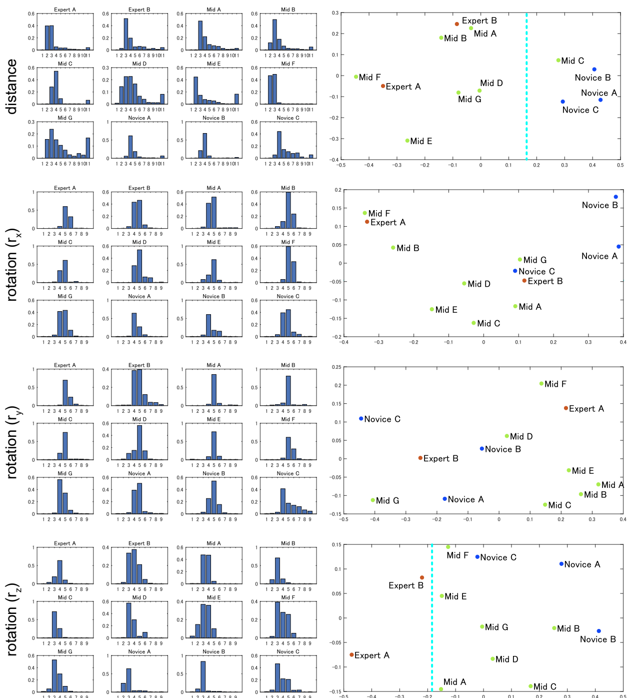

Figure 7. PCA anaylis of the Experiment 1. (Left) Histograms of mutual facial distances and poses (r_x, r_y and r_z). (Right) PCA score results where the x and y axes show the first and second components. There are clear thresholds between novices and others at about $x = 0.16$ for the distance PCA score, and between experts and others at about $x = -0.18$ in r_z PCA score.

of the care receiver's face in the FPV image plane (the plane perpendicular to the facial frontal direction) as shown in Figure 4. Namely, the average and peak z-rotation values of the experts and the middle-level caregivers were located around 0 [deg] while those of novices were much larger.

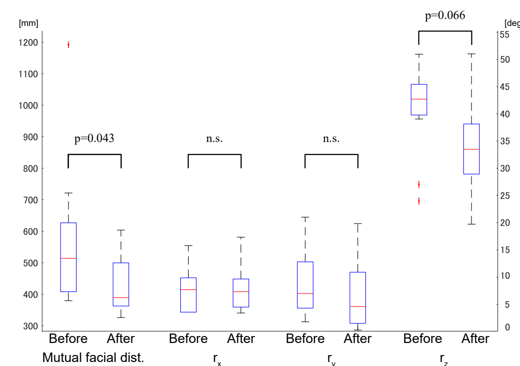

Figure 8. The result of Experiment 2. Differences of average mutual facial distances and poses before and after taking Humanitude learning sessions of 12 novices.

5.2. Experiment 2: Skill progressions while face-to-face communication before and after taking care-learning sessions

In the second experiment, we compared the mutual facial distance and pose of novice caregivers before and after taking Humanitude care sessions. We obtained the number of facial distances and poses from the dataset of 4.2 and find the difference.

Analysis and results: The average mutual facial distance and poses (angles) are shown in Table 2 and Figure 8. We could find the significant difference in the average mutual facial distance between caregiver and receivers ($p = 0.043$), and marginally significance in r_z ($p = 0.066$), which follows the same trend as the Experiment 1.

Thus, we further performed the PCA analysis for the 20-dimensional feature vector combining the distance ($\mathbf{f}_{\text{dist}}$) and z-angular ($\mathbf{f}_{r_z}$) histograms. Figure 9 shows the result. In here, we first obtain the PCA subspace of the data in the experiment 1, then map all the data in the subspace. As the result, we can clearly observe 1) the clusters of novices, middle level, novices after taking the Humanitude care ses-

Table 2. The result of Experiment 2.

Name	Before training						After training					
	Total frames	Face detected	Av. mutual facial dist.	Av. mutual facial pose rx	ry	rz	Total frames	Face detected	Av. mutual facial dist.	Av. mutual facial pose rx	ry	rz
Novice D	3932	1656 (42.1%)	1191.61	-3.19	-10.64	26.82	1889	993 (52.6%)	364.32	10.00	-5.16	45.99
Novice E	1154	540 (46.8%)	404.16	3.54	-2.34	43.94	3000	876 (29.2%)	436.96	3.35	-13.52	33.49
Novice F	2993	1341 (44.8%)	721.4	5.79	-1.59	42.36	1441	781 (54.2%)	406.4	8.91	0.08	36.18
Novice G	1376	574 (41.7%)	427.46	8.17	-2.30	38.85	2307	698 (30.3%)	326.16	6.91	-1.54	50.82
Novice H	624	237 (38.0%)	411.53	7.71	-5.87	44.01	1275	475 (37.3%)	368.73	8.62	-3.64	22.34
Novice I	971	345 (35.5%)	404.63	1.19	-20.82	55.17	2932	1134 (38.7%)	372.2	3.35	-9.61	32.22
Novice J	4704	2116 (45.0%)	535.36	7.27	-7.05	46.44	4523	1082 (23.9%)	476.04	15.44	-9.02	35.55
Novice K	1773	714 (40.3%)	379.49	15.59	9.79	23.77	1518	427 (28.1%)	523.26	7.38	1.07	19.53
Novice L	804	345 (42.9%)	697.66	11.27	-14.57	50.76	1071	403 (37.6%)	526.74	6.51	-19.64	33.10
Novice M	920	563 (61.2%)	555.78	10.28	-6.52	40.31	3816	2443 (64.0%)	332.86	3.21	-0.65	39.70
Novice N	746	179 (24.0%)	529.79	9.06	-15.59	41.39	2550	1442 (56.5%)	361.32	17.14	-1.81	25.56
Novice O	819	315 (38.5%)	498.8	-0.12	-6.06	42.65	1462	630 (43.1%)	603.64	5.24	-11.80	31.91

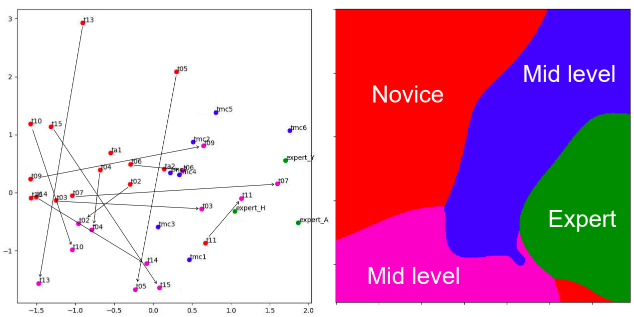

Figure 9. (Left) PCA analysis of all dataset. We performed the PCA for 20-dimensional feature vectors ($\mathbf{f}_{\text{dist}}$ and $\mathbf{f}_{r_z}$) and plot the PCA scores of the data. In here, x-axis and y-axis denote the 1st and 2nd components of PCA score. The green markers indicate experts, blue markers indicate middle-levels, purple markers indicate the novices after taking the Humanitude care learning sessions, and red markers indicate the novices before taking the Humanitude care learning session. We can clearly observe the contiguous clusters from novices, middle levels and experts in the space. The allows shows the relation of the data before and after taking the care learning session, (right) visualization of expert, middle and novice clusters.

sion, and experts, and 2) the clusters are contiguous in the PCA space. This finding shows that the feature vector used for the analysis reflects the skill level of the tender-care, and the potential to evaluate the skill level of the new learners.

According to the chronological difference of the behavioral data before and after taking the care sessions, eight participants got the higher scores of the 1st component of PCA subspace, while four participants decreased the score. Since the experts' scores of the 1st component are at the positive end, we think the 1st component of the PCA corresponds to the face-to-face skill in the Humanitude.

6. Discussion and conclusion

Unsupervised analysis results of mutual facial distances and facial poses enabled us to find significant differences between novices, middle-level and expert Humanitude caregivers. Specially, we found a clear threshold in eye contact frequency and PCA scores of facial distance and r_z-rotation histograms, which indicate that the important skills in Humanitude tender-care are related to a) frequent eye contact, b) a nearest mutual facial distance of less than 30 [cm] and c) mutual facial poses being in the same direction. This can also be seen from the 1st PCA components of the distance and r_z histograms.

In the experiment 2, we applied the same analysis to observe the skill progressions of the novice care learners. As the result, we found the clear contiguous clusters of novices, middle levels and experts in the PCA subspace. We also found that 8 out of 12 novices obtain the larger scores in the 1st component of the PCA, which means their behaviors become closer to the experts. These skills are a part of Humanitude gaze skill: caregivers should communicate to the care receivers while keeping eye contact from a close distance and possessing the same facial angles of the care receiver's face. This is based on the idea of Humanitude care methodology that all behaviors are considered to imply nonverbal messages. To have the eye contact straight in front of the care receiver expresses the fairness, and the distance between caregivers and care receivers reflects their friendliness. The study results show that the experts expressed fairness and friendliness much more than the novices. This skill is a core skill with which to establish a good relationship that leads to high-quality care.

Acknowledgements: All the experiments were conducted in compliance with the protocol which was reviewed and approved by the ethical committee of Unit for Advanced Studies of the Human Mind, Kyoto University (Permit Number: 30-P-4). This work was supported by JST CREST Grant Number JPMJCR17A5 and JSPS KAKENHI 17H01779, Japan.

References

[1] G. J. Acton and J. Kang. Interventions to reduce the burden of caregiving for an adult with dementia: A meta-analysis.

Research in nursing & health, 24(5):349–360, 2001. 1

[2] R. D. Adelman, L. L. Tmanova, D. Delgado, S. Dion, and M. S. Lachs. Caregiver burden: a clinical review. *Jama*, 311(10):1052–1060, 2014. 1

[3] Amazon Inc. Amazon rekognition, 2019. 3

[4] T. Baltrusaitis, P. Robinson, and L. P. Morency. OpenFace: An open source facial behavior analysis toolkit. *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016*, 2016. 3

[5] G. Bertasius, H. S. Park, X. Y. Stella, and J. Shi. Am i a baller? basketball performance assessment from first-person videos. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pages 2196–2204. IEEE, 2017. 2

[6] S. Biquand and B. Zittel. Care giving and nursing, work conditions and humanitude®. *Work*, 41(Supplement 1):1828–1831, 2012. 1

[7] S. Boseley. Dementia research funding to more than double to £66m by 2015. *The Guardian*, 2012. 1

[8] L. C. Campbell, F. J. Keefe, C. Scipio, D. C. McKee, C. L. Edwards, S. H. Herman, L. E. Johnson, O. M. Colvin, C. M. McBride, and C. Donatucci. Facilitating research participation and improving quality of life for african american prostate cancer survivors and their intimate partners: a pilot study of telephone-based coping skills training. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 109(S2):414–424, 2007. 1

[9] B. Casado and P. Sacco. Correlates of caregiver burden among family caregivers of older korean americans. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 67(3):331–336, 2011. 1

[10] L. D. Clyburn, M. J. Stones, T. Hadjistavropoulos, H. Tuokko, et al. Predicting caregiver burden and depression in alzheimer's disease. *Journals of Gerontology series b*, 55(1):S2–S13, 2000. 1

[11] D. W. Coon, L. Thompson, A. Steffen, K. Sorocco, and D. Gallagher-Thompson. Anger and depression management: psychoeducational skill training interventions for women caregivers of a relative with dementia. *The Gerontologist*, 43(5):678–689, 2003. 1

[12] J. J. Dunkin and C. Anderson-Hanley. Dementia caregiver burden: a review of the literature and guidelines for assessment and intervention. *Neurology*, 51(1 Suppl 1):S53–S60, 1998. 1

[13] L. Etters, D. Goodall, and B. E. Harrison. Caregiver burden among dementia patient caregivers: a review of the literature. *Journal of the American Academy of Nurse Practitioners*, 20(8):423–428, 2008. 1

[14] A. Fathi, J. K. Hodgins, and J. M. Rehg. Social interactions: A first-person perspective. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1226–1233. IEEE, 2012. 2

[15] Y. Gineste and J. Pellissier. *Humanitude: comprendre la vieillesse, prendre soin des hommes vieux*. A. Colin, 2007. 2

[16] B. Given, P. R. Sherwood, and C. W. Given. What knowledge and skills do caregivers need? *Journal of Social Work Education*, 44(sup3):115–123, 2008. 1

[17] M. Honda, M. Ito, S. Ishikawa, Y. Takebayashi, and L. Tierney. Reduction of behavioral psychological symptoms of dementia by multimodal comprehensive care for vulnerable geriatric patients in an acute care hospital: A case series. *Case reports in medicine*, 2016, 2016. 2

[18] S. Ishikawa, M. Ito, M. Honda, and Y. Takebayashi. The skill representation of a multimodal communication care method for people with dementia. *JJAP Conference Proceedings*, 011616:4, 2016. 2

[19] H. Kim, M. Chang, K. Rose, and S. Kim. Predictors of caregiver burden in caregivers of individuals with dementia. *Journal of advanced nursing*, 68(4):846–855, 2012. 1

[20] E. B. Larson, K. Yaffe, and K. M. Langa. New insights into the dementia epidemic. *New England Journal of Medicine*, 369(24):2275–2277, 2013. 1

[21] H. Law, K. Ghani, and J. Deng. Surgeon technical skill assessment using computer vision based analysis. In *Machine Learning for Healthcare Conference*, pages 88–99, 2017. 2

[22] M. Ma, H. Fan, and K. M. Kitani. Going deeper into first-person activity recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1894–1903, 2016. 2

[23] A. Nakazawa, Y. Mitsuzumi, Y. Watanabe, R. Kurazume, S. Yoshikawa, and M. Honda. First-person video analysis for evaluating skill level in the humanitude tender-care technique. *Journal of Intelligent & Robotic Systems*, Jul 2019. 2, 3

[24] L. L. Northouse, M. C. Katapodi, L. Song, L. Zhang, and D. W. Mood. Interventions with family caregivers of cancer patients: meta-analysis of randomized trials. *CA: a cancer journal for clinicians*, 60(5):317–339, 2010. 1

[25] S. K. Ostwald, K. W. Hepburn, W. Caron, T. Burns, and R. Mantell. Reducing caregiver burden: a randomized psychoeducational intervention for caregivers of persons with dementia. *The Gerontologist*, 39(3):299–309, 1999. 1

[26] E. Papastavrou, A. Kalokerinou, S. S. Papacostas, H. Tsangari, and P. Sourtzi. Caring for a relative with dementia: family caregiver burden. *Journal of advanced nursing*, 58(5):446–457, 2007. 1

[27] Pupil Labs. Pupil labs camera system, 2018. [Online; accessed 25-Aug-2018]. 2

[28] S. Singh, C. Arora, and C. Jawahar. First person action recognition using deep learned descriptors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2620–2628, 2016. 2

[29] H. Soo Park, J. Shi, et al. Force from motion: decoding physical sensation in a first person video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3834–3842, 2016. 2

[30] S. Su, J. P. Hong, J. Shi, and H. S. Park. Predicting behaviors of basketball players from first person videos. In *CVPR*, volume 2, page 3, 2017. 2

[31] K. K. Win, M. S. Chong, N. Ali, M. Chan, and W. S. Lim. Burden among family caregivers of dementia in the oldest-old: an exploratory study. *Frontiers in medicine*, 4:205, 2017. 1

[32] World Health Organization et al. Dementia: Fact sheet N 362, 2012. 1

[33] Z. Ye, Y. Li, A. Fathi, Y. Han, A. Rozga, G. D. Abowd, and J. M. Rehg. Detecting eye contact using wearable eye-tracking glasses. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, pages 699–704, New York, NY, USA, 2012. ACM. 3