

STARR: Decision Support and self-mAnagement system for stRoke survivoRs Vision based Rehabilitation System

Abd El Rahman Shabayek¹, Renato Baptista¹, Konstantinos Papadopoulos¹,
Girum Demisse¹, Oyeade Oyedotun¹, Michel Antunes, Djamila Aouada¹,
Bjorn Ottersten¹, Margarita Anastassova², Mehdi Boukallel², Sabrina Panëels²,
Gary Randall³, Mathilde Andre⁴, Alice Douchet⁴, Stephane Bouilland⁴
and Leire Ortiz Fernandez⁵

¹Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg,
Luxembourg

²CEA, LIST, Sensory and Ambient Interfaces Laboratory, France

³Stroke Association, United Kingdom

⁴Fondation Hopale, France

⁵Biocruces Health Research Institute, Cruces University Hospital, Osakidetza, Spain

{AbdElRahman.Shabayek, Renato.Baptista,
Konstantinos.Papadopoulos, Girum.Demisse, Oyeade.Oyedotun,
Djamila.Aouada, Bjorn.Ottersten}@uni.lu

Gary.Randall@stroke.org.uk

{Margarita.Anastassova, Mehdi.Boukallel, Sabrina.Paneels}@cea.fr

{Mathilde.Andre, Alice.Douchet, Stephane.Bouilland}@hopale.com

Leire.Ortizfernandez@osakidetza.eus

Abstract. This chapter explains a vision based platform developed within a European project on decision support and self-management for stroke survivors. The objective is to provide a low cost home rehabilitation system. Our main concern is to maintain the patients' physical activity while carrying a continuous monitoring of his physical and emotional state. This is essential for recovering some autonomy in daily life activities and preventing a second damaging stroke. Post-stroke patients are initially subject to physical therapy under the supervision of a health professional to follow up on their daily physical activity and monitor their emotional state. However, due to social and economical constraints, home based rehabilitation is eventually suggested. Our vision platform paves the way towards having low cost home rehabilitation.

1 Introduction

According to the World Health Organization (WHO), 17 million people worldwide suffer a stroke each year [1]. Currently, rates of deaths from stroke in different countries range from 30 per 100,000 of the population to 170 per 100,000 of the population. Even though preventive measures (e.g., control of blood pressure, cholesterol checks, etc.) are in place by national health systems to reduce the incidence of stroke, the number of persons having a stroke in Europe is likely to increase. The projections in [2] indicate that

between 2015 and 2035, there will be a 34% increase in total number of stroke events in the EU from 613,148 in 2015 to 819,771 in 2035 solely because of the demographic changes related to an increasing aging population.

Ongoing improvements in Stroke treatment lead to greatly reduced death rates. Hence, there are more people surviving their strokes and living with the consequences. So the 45 billion euros estimated cost of stroke in Europe (healthcare and non-healthcare costs) in 2015 is expected to increase [2].

Recurrent or secondary stroke carries with it a greater risk than first-ever stroke for death and disability [3, 4]. As mortality from first time stroke has decreased over the past years, the number of people at risk for a secondary stroke has increased, with an associated increase in healthcare costs [5]. In the same time, secondary stroke prevention has proved successful in experimental trials [6].

In stroke, as in any other chronic condition, patients self-manage their behaviors [7]. Hence, it is crucial to provide stroke survivors with continuous support to help them better manage their condition, live independently, and prevent a secondary stroke [8]. This is the primary objective of the Decision SupportT and self-mAnagement system for stRoKe survivoRs (STARR) project [9]. STARR aims to develop a modular, affordable, and easy-to-use, install and maintain system, which will inform stroke survivors about the relation between their daily activities and the risk of having a secondary stroke. See Fig. 1 for a synthetic view of the STARR system. A vision-based platform to support rehabilitation of stroke survivors is developed as part of the self-management services. This Chapter describes the different algorithms and concepts used in building the vision-based platform of the STARR system. These algorithms are related to the analysis and recognition of human 1) physical activity, which is addressed in Section 2 and 2) emotions, addressed in Section 3. The Chapter is concluded and future work is proposed in Section 4.

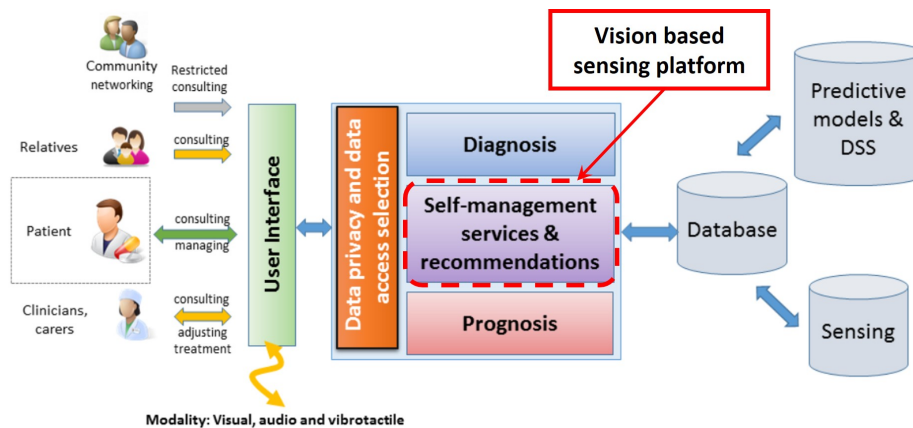


Fig. 1. A synthetic view of the STARR system. The vision based platform is part of the services provided by the "Self-management services & recommendations" component in the STARR system [9].

2 Human Motion Analysis

It is critical for stroke survivors that they recover some autonomy in their daily life activities [10]. Despite the benefits of physical activity, many stroke survivors do not exercise regularly due to many reasons, such as fatigue, lack of motivation, confidence, and skill levels [11]. Traditionally, post-stroke patients are initially subject to physical therapy under the supervision of a health professional aimed at restoring and maintaining activities of daily living in rehabilitation centers [12], and continuously advises her/him in how to improve the motion as well as interrupts the exercise in case of health related risk issues. In chronic phase, patients must continue doing exercises and recommendations at home.

We developed algorithms for guiding patients to perform correctly an action or movement by presenting feedback proposals in the form of visual information and human-interpretable messages. Our proposed models use a 3D human body skeleton representation [18–20].

This section is divided into two parts: 1) feedback proposals for assisting human physical activity where both static and dynamic actions are considered, and 2) human action recognition where complete and incomplete (i.e. early detection) actions are considered.

2.1 Feedback for Assisting Human Physical Activity

Feedback on Static Human Actions. The core of the proposed approach [18] is the analysis of the motion required for aligning corresponding body-parts with respect to a template skeleton pose. The concept of the system is based on the creation of specific static template poses (i.e. these template poses can be created by a physiotherapist as a prescription for the patient to exercise) as a reference such that the patient tries to match with the guidance of automatic and human-interpretable feedback proposals.

The algorithm starts by performing a spatial and temporal alignment of the skeleton data of the patient with respect to the template pose. Then, the objective is to compute the motion that each body-part of the patient needs to undergo to better match the template pose. Note that the human skeleton information is decomposed into a predefined set of body-parts and there is no learning involved. The computation of per-body-part motion can be found in a closed form solution.

The objective of this work was not only to assess the quality of an action, but also to provide feedback in how to improve the movement being performed. We do not compute feedback for single joints, but for body-parts, defined as configurations of skeleton joints that may or may not move rigidly. Feedback proposals are automatically computed by comparing the movement being performed with a template pose, without specifying pose constraints of joint configurations. Feedback instructions are not only presented in the form of visual information, but also human-interpretable feedback is proposed from discretized spatial transformations. Figure 2 shows an example of how the feedback proposals are provided to the patient.

Feedback on Dynamic Human Actions. We explored the concept of providing feedback to a user moving in front of a depth camera so that (s)he is able to replicate a

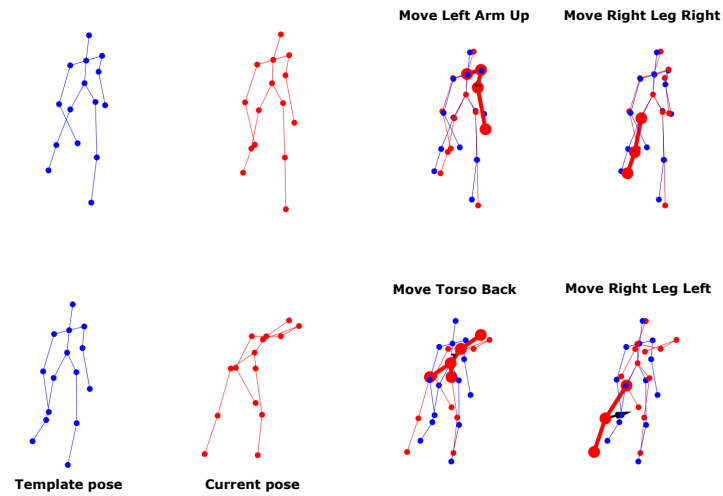


Fig. 2. Feedback proposals in the form of visual information (highlighted body-part and an arrow, anchored to the body-part centroid, pointing to the direction of the motion) and human-interpretable messages (e.g., “Move Left Arm Up” or “Move Torso Back”). In the first row, an example of normal people walking is shown, and in the second row, an example of a walking stroke survivor is shown. Figure adapted from [18].

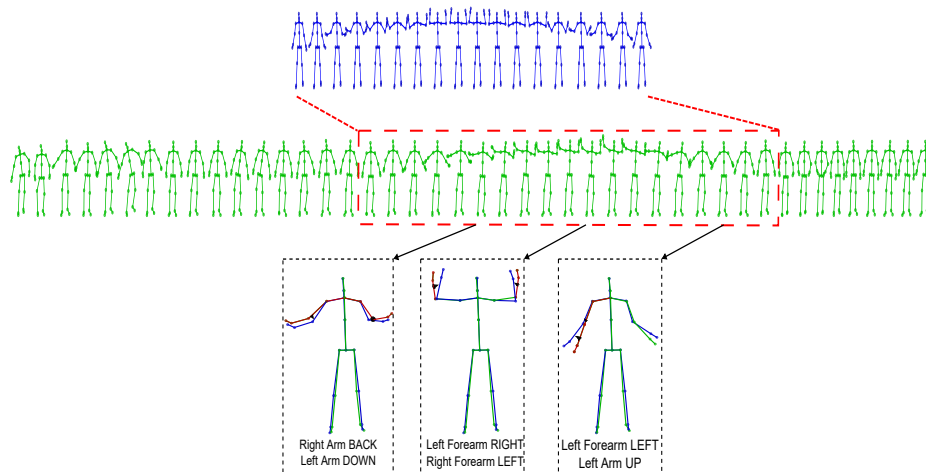


Fig. 3. Temporal alignment using SS-DTW and posterior feedback proposals. The first row represents the template action (blue color sequence), the second row (green color sequence) is the performed action by the user. The interval corresponding to the temporal alignment is represented by the red rectangle. Feedback proposals are shown for the same instance for both methods in order to correct the position with respect to the template action. Figure adapted from [19].

specific template action [19]. Note that, we denote a template action as a dynamic template action (i.e series of poses) to differ from the previous section, where the static template (i.e a single pose) was considered.

In the dynamic case, the user is guided in how to correctly perform a certain template action by following feedback proposals while exercising. These proposals are presented in the form of visual information and in a human-interpretable manner similar to [18]. In order to align an action that was performed with the template action, we explored two different approaches, namely, Subsequence Dynamic Time Warping (SS-DTW) [26] and Temporal Commonality Discovery (TCD) [27]. The first method aims to find the optimal temporal alignment and the second one discovers the interval where both sequences share similar content, after which standard Dynamic Time Warping (DTW) is applied for the temporal alignment. Additionally, feedback proposals are provided in order to correct the user with respect to the template action. Experimental results show that both methods have similar accuracy rate and the computational time is a decisive factor, where Subsequence Dynamic Time Warping achieves faster results. Figure 3 shows the results of the SS-DTW and the feedback proposals provided to correct the user for each time instance.

Feedback on Human Posture. The objective is to make sure that while exercising, the patient shall keep a correct healthy posture. A correct posture is defined as a position in which minimum stress is applied to each body joint [28]. Hence, we proposed a guidance framework to correct the patient's posture in real-time without requiring a physical or a direct intervention of a therapist [20].

The built system evaluates continuously postural defects of the patient. If there are deviations from a correct posture, then feedback information is provided in order to guide the user to converge to an appropriate and stable body condition. The core of the proposed approach is the analysis of the motion required for aligning body-parts with respect to postural constraints. For a correct posture, we focused on two aspects to define a correct posture. The first is related to having a straight back that is aligned with the gravity vector, see Figure 4(a), and the second aspect refers to the balance between left and right limbs, see Figure 4(b). In other words, the body joints of the limbs should be symmetric about the plane of symmetry that intersects the line of gravity.

Posture analysis is also very important when it comes to the assessment of sports movements. For example, lifting a weight may bring spine injuries in a long term if not performed correctly. Most people incorrectly lift a weight by bending and executing most of the force using their back. While lifting a weight, the ideal body posture is having a balance between the left and right limbs and also having the back as straight as possible. In this case, most of the force is applied on the leg muscles. Figure 5(a) shows an example of a correct versus an incorrect lifting weight body posture. Figure 5(b) shows the angle variation between the back and the gravity vector while lifting a weight.

2.2 Human Action Recognition

Action recognition strongly supports home-based rehabilitation, reducing the need of having a physiotherapist supervising the patient. However, it remains a challenging task

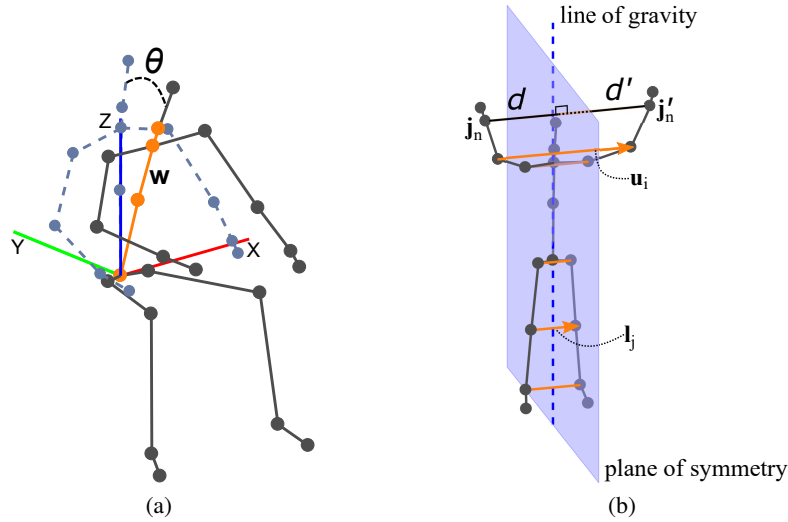


Fig. 4. Figure 4(a) illustrates the angle between the spine vector (vector represented in orange color) and the gravity vector (aligned with the z -axis - in blue color). In Figure 4(b) the purple plane represents the plane of symmetry, which divides the human skeleton into two parts (left and right sides of the human skeleton). Both figures were adapted from [20].

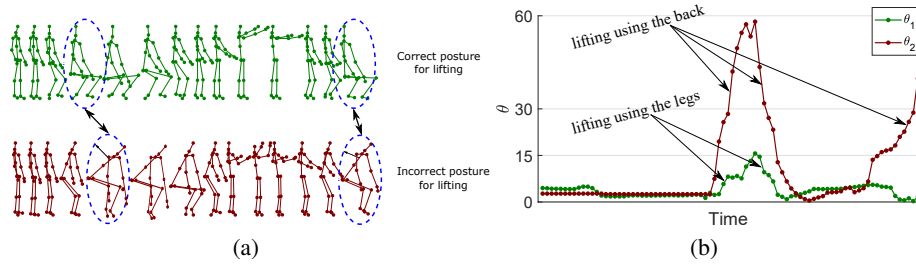


Fig. 5. Figure 5(a) shows an example of the lifting exercise. The green color sequence represents the exercise maintaining a correct posture, and the red color sequence refers to the incorrect posture. Figure 5(b) plots the variation of the angle between the back and the gravity vector over time. Both figures were adapted from [20].

due to large intra-class variation, strong view-dependency and occlusions. We have worked on both complete and incomplete action recognition to support different home scenarios done by the patient. These different techniques will be highlighted in the following sections.

Detection and Recognition of Complete Actions. One of our main objectives is to provide intelligent assistive system using action recognition for stroke survivors. Patients' movements can be described using dense motion trajectories. These trajectories represent the motion as a vector of point displacement on the image grid. We proposed a human-tailored trajectory extraction scheme [21], in which similar motion trajectories are clustered around specific body joints using information relevant to the human pose.

This approach offers doctors and therapists the ability to track and isolate the dense motion fields of each body joint, focusing on the ones which have the biggest contribution to a specific movement.

We developed two approaches; in the first approach, we made the assumption that skeleton joint positions are provided, whereas in the second approach, only an estimate of the joints is available. In both cases, local Bag-of-Words concept is further empowering the proposed methods, generating a specific codebook for each skeleton joint cluster. The final video representation is benefited by the direct addition of spatial awareness of human pose, effectively increasing its discriminative power and rejecting non-informative motion trajectories that could confuse the recognition process. Figure 6 shows a comparison between conventional single BoW and multiple local BoW action recognition approaches.

In addition to the action recognition concept described above, we conducted a study on trajectory-based action detection. Our goal was to use this concept in a potent and non-uniform way for detecting the starting and ending points of patient's specific movements. In general, improved trajectories have been proven to be effective for motion description in action recognition, but they are not sufficiently exploited on action detection.

We proposed a two-step enhanced approach making use of the improved trajectories [22]. This approach allows the implementation of an independent home-based rehabilitation system without the need of a therapist's supervision. The first step is the detection of action proposals (i.e possible candidate actions) and takes place in an online manner. In a sequence of continuous activities performed by the patient, this approach isolates specific action proposals by labeling each frame as a positive or negative action. As soon as the action proposals are generated, the second step, which is the classification using improved trajectories, is performed. Figure 7 shows the proposed enhanced model for action detection.

Recognition of Incomplete Actions. In certain scenarios, e.g. fall detection, early action recognition is required. We proposed to detect an action as soon as possible and ideally before it is fully completed [23]. For such a scenario, it is of significant importance to have low detection time latency rather than having high detection and recognition rates. Our solution consists in an adaptive sliding window approach, which efficiently rejects irrelevant data during streaming. For an added efficiency, we proposed to use partial information from the template actions. Hence, we presented an analysis of the influence of using partial information of the action to be detected. With this analysis, we were able to achieve competitive detection results when compared to using the complete information of the action template. Figure 8 shows an example of the skeleton representation of the human body while performing a suspicious action and the corresponding detection.

3 Human Emotion Analysis

Following a stroke, survivors suffer from a large distress expressed through post-stroke depression (PSD), post-stroke anxiety, post-stroke emotional incontinence (PSEI), post-

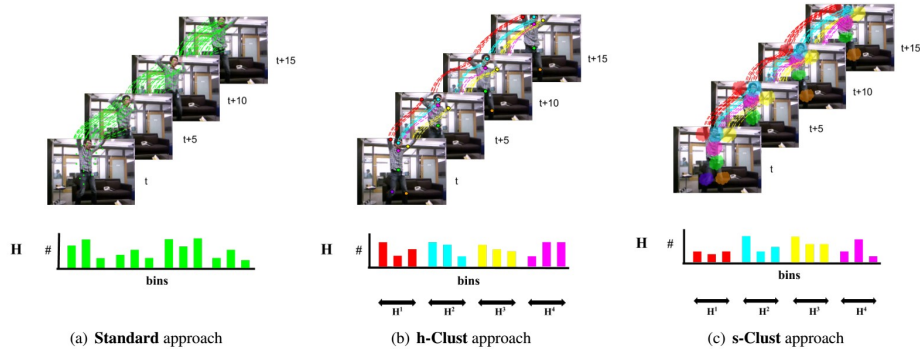


Fig. 6. Standard, H(ard)-Clust(ering) and S(oft)-Clust(ering) approaches. Left: in the standard approach a single BoW is computed from the entire set of the extracted motion trajectories; Middle: motion trajectories are uniquely assigned to the corresponding skeleton joint cluster, and a local BoW is generated for each one of them. Right: motion trajectories are assigned to the corresponding skeleton joint cluster using a soft-voting scheme applied on the joint location heatmaps; the utilization of local BoW remains the same. Top: clustered dense trajectories, where different colors identify different clusters. Bottom: histogram of visual words. Adapted from [21].

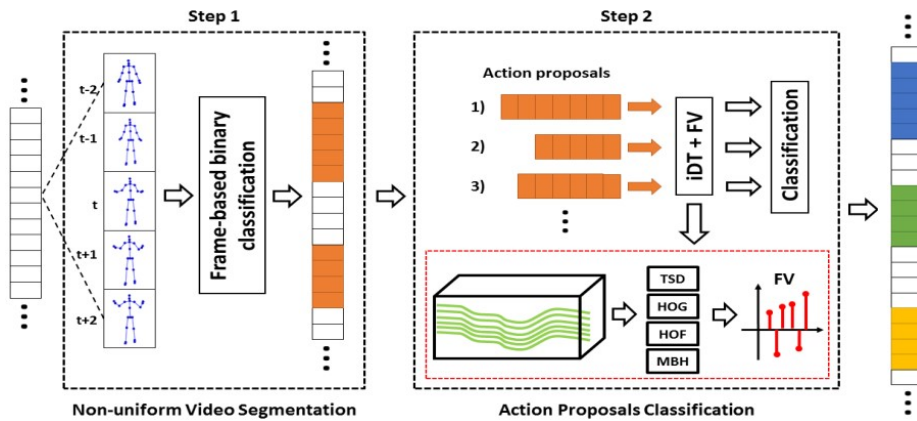


Fig. 7. Trajectory-based action detection approach. In Step 1, skeleton joint features from a temporal window around the current frame are used as input to a classifier for the generation of action proposals from the input sequence. In Step 2, improved trajectories action recognition is performed on the resulting action proposals. Adapted from [22].

stroke anger proneness (PSAP), and post-stroke fatigue [14–16]. Their emotional well-being is important throughout, especially when interacting with rehabilitation systems at home such as our vision based platform proposed in STARR. It would therefore be very informative for therapists to know how patients are feeling during an exercise. The session could then be adapted to the users feelings and engagement with the system by, for example, increasing motivational messages, changing the level of exercising difficulty, or changing the exercising scenario. However, there is still existing challenges as facial expressions may not strongly indicate the overall mood or motivation levels. Even if they were strongly predictive, the measured signals might be weak and brief.

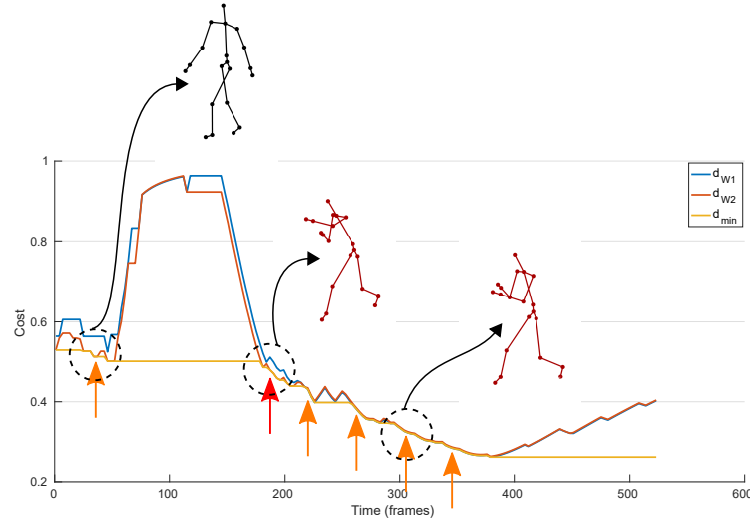


Fig. 8. Skeleton representation of the human body while performing a suspicious action, in this case: Punching. The orange arrows represent the alarms that were generated over time and the red arrow is the alarm of interest which is the first alarm generated within the ground truth. Figure adapted from [23].

Another serious challenge is that many strokes in the sub-acute phase will have facial paralysis, which, at the least, is a compounding factor.

In the following sections we highlight our work on emotion analysis, namely on model based and learning based 3D facial expression representations.

3.1 Model based 3D Facial Expression Representation

We proposed a deformation based representation for analyzing expressions from 3D faces [24]. Given a fixed starting point, a 3D facial point cloud is decomposed into an ordered deformable set of curves. Subsequently, the set of curves are represented by an element of a high dimensional matrix Lie group, specifically the direct product of $SE(3)$. The representation has two main advantages. First, facial expressions can be decoupled from a neutral face using the group structure of the representation. This is particularly important to model facial deformations, exclusively. Second, an underlying non-linear facial expression manifold can be linearized for further processing. This opens up the possibility of using conventional statistical and linear models for facial expression analysis without compromising the underlying manifold. Furthermore, linear combinations of facial expressions can be simulated by using the group and differential structure of the representation. Figure 9 shows our proposed model.

3.2 Learning Based 3D Facial Expression Representation

Deep convolutional neural networks (DCNN) show very good results on regular 2D data of faces. Given that we have, in addition, 3D data captured simultaneously with an

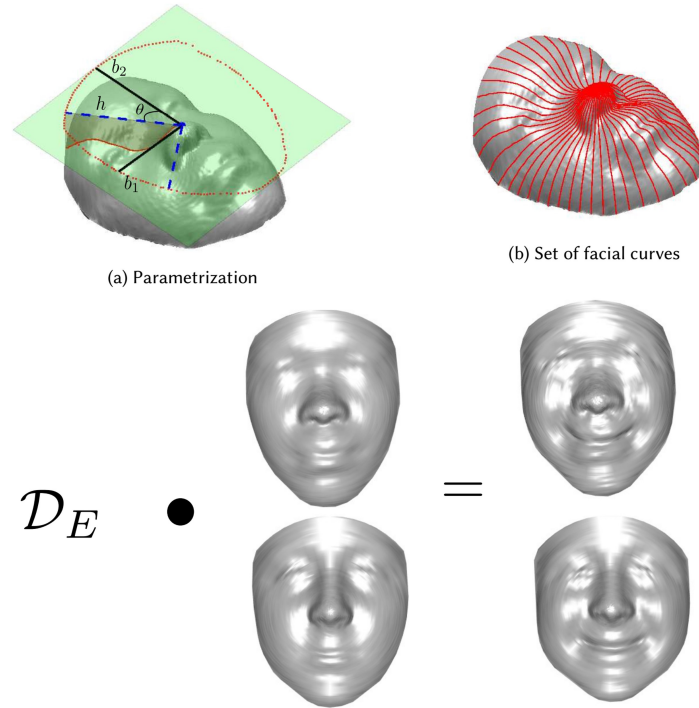


Fig. 9. Top row shows facial curve extraction. The bottom row shows the proposed 3D facial expression representation. A smile is captured by the deformation \mathcal{D}_E and applied to two different subjects, preserving their specific shapes. Adapted from [24].

RGB-D (e.g. Kinect) sensor, it is natural to consider further improving the performance of emotion sensing from 2D facial data. We proposed to use DCNNs for joint learning of robust facial expression features from fused 2D and 3D data [25]. We used the pre-trained residual network architecture ResNet50 [17] for learning features from 2D data, while we train a DCNN from scratch on the 3D data. Our proposed fusion approach outperforms learning singly from either 2D or 3D data. The way features are learned from 3D data is important. The simplest approach, is to represent 3D data with depth maps. A more elaborate representation of 3D features could make a difference in the final performance.

4 Conclusion and Future Work

In this chapter, we overviewed the different computer vision algorithms proposed as building components of a vision based platform for stroke survivors home rehabilitation as part of the European project STARR. Our research efforts are directed in two folds: human physical activity analysis and emotion analysis. Human physical activity is analyzed to propose human-interpretable feedback on both static and dynamic actions combined with real-time posture analysis. Furthermore, complete and incomplete action recognition approaches were proposed to strongly support patient activity mo-

monitoring at home. In order to monitor the patients emotional state, model based and learning based facial emotion analysis approaches were explored and new approaches were proposed.

Our future research direction with respect to physical activity is toward analyzing its level of fluidity. As the movement quality for stroke survivors is characterized by it. In fact, a more fluid movement is one that is less spastic. Using the same skeleton model of the human body, we attempt to detect spasticity when it happens by clustering the signatures of each motion into the corresponding modified Ashworth scale [13]. This is important in order to detect spasticity immediately as it happens and act accordingly by stopping or re-adjusting the system. It is also important for doctors to monitor the mid-term progress of a patient and detect as early as possible any signs of relapse.

A software package will be developed with two applications; one for the therapist to prescribe the exercise and another for the patient to load it. Our algorithms will function on both sides in order to enable the exercise prescription and to provide the required analysis.

Acknowledgement. This work was funded by the European Unions Horizon 2020 research and innovation project STARR under grant agreement No.689947, and partly supported by the National Research Fund (FNR), Luxembourg, under the projects C15/IS/10415355/3DACT/Bjorn Ottersten and R-AGR-0424-05-D/Bjorn Ottersten.

References

1. World Health Organization. The top 10 causes of death. January 2017 [cited 13/03/17];Factsheet].
2. King's College London for The Stroke Alliance for Europe (SAFE): The Burden of Stroke in EUROPE. Report. <http://www.strokeeurope.eu>. (2018)
3. Samsa, G.P., Bian, J., Lipscomb, J., Matchar, D.B.: Epidemiology of recurrent cerebral infarction: A medicare claims-based comparison of first and recurrent strokes on 2-year survival and cost. *Stroke*. (1999) 30:338-349
4. Jorgensen, H.S., Nakayama, H., Reith, J., Raaschou, H.O., Olsen, T.S.: Stroke recurrence: Predictors, severity, and prognosis. The Copenhagen stroke study. *Neurology*, (1997) 48:891-895.
5. Williams, G.R.: Incidence and characteristics of total stroke in the united states. *BMC Neurology*. (2001) 1:2
6. Qureshi, A.I., Suri, M.F., Guterman, L.R., Hopkins, L.N.: Ineffective secondary prevention in survivors of cardiovascular events in the US population: Report from the third national health and nutrition examination survey. *Archives of Internal Medicine*. (2001) 161:1621-1628
7. Bodenheimer, T., Lorig, K., Holman, H., Grumbach, K.: Patient self-management of chronic disease in primary care. *Journal of the American Medical Association*, (2002) 288: 2469-2475
8. Newman, S., Steed, L., Mulligan, K.: Self-management interventions for chronic illness. *Lancet*. (2004) 364: 1523-1537
9. <http://www.starrproject.org/>
10. Kwakkel, G., Kollen, B.J., Krebs, H.I.: Effects of robot-assisted therapy on upper limb recovery after stroke: a systematic review. *Neurorehabilitation and neural repair*.

11. Oi, F., Kurillo, G., Obdrzalek, S., Bajcsy, R., Jimison, H.B., Pavel, M.: Design and evaluation of an interactive exercise coaching system for older adults: Lessons learned. *IEEE J. Biomedical and Health Informatics*. (2016)
12. Veerbeek, J.M., van Wegen, E., van Peppen, R., van der Wees, P.J., Hendriks, E., Rietberg, M., Kwakkel, G.: What is the evidence for physical therapy poststroke? a systematic review and meta-analysis. *PloS one*. (2014)
13. Ansari, NN; Naghdi, S; Arab, TK; Jalaie, S: The interrater and intrarater reliability of the Modified Ashworth Scale in the assessment of muscle spasticity: limb and muscle group effect. *NeuroRehabilitation*. (2008) 23 (3): 2317.
14. Gillespie, C., Joice S., Lawrence, M., Whittick, J.: Interventions for post-stroke disturbances of mood and emotional behaviour: A review of SIGN 118 stroke rehabilitation guideline recommendations. *Int J Ther Rehabil*. (2011) 545-553
15. Gillespie, C., Cadden, A.: The Distress Management System for Stroke (DMSS): An approach for screening and initial intervention for post-stroke psychological distress. *Journal of Nursing Education and Practice*. (2013)
16. Kneebone, I., Neffgen, L., Pettyfer, S.: Screening for depression and anxiety after stroke: developing protocols for use in the community. *Disability and rehabilitation*. (2012) 1114-1120
17. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition*. (2016).
18. Antunes, M., Baptista, R., Demisse, G., Aouada, D., Ottersten, B.: Visual and human-interpretable feedback for assisting physical activity. *European Conference on Computer Vision (ECCV) Workshop on Assistive Computer Vision and Robotics Amsterdam*. (2016).
19. Baptista, R., Antunes, M., Aouada, D., Ottersten, B.: Video-Based Feedback for Assisting Physical Activity. *12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP)*. (2017)
20. Baptista, R., Antunes, M., Shabayek, AER., Aouada, D., Ottersten, B.: Flexible Feedback System for Posture Monitoring and Correction. *IEEE International Conference on Image Information Processing (ICIIP)*. (2017)
21. Papadopoulos, K., Antunes, M., Aouada, D., Ottersten, B.: Enhanced Trajectory-based Action Recognition using Human Pose. *IEEE International Conference on Image Processing, Beijing, China*. (2017)
22. Papadopoulos, K., Antunes, M., Aouada, D., Ottersten, B.: A Revisit of Action Detection using Improved Trajectories. *IEEE International Conference on Acoustics, Speech and Signal Processing, Calgary, Alberta, Canada*. (2018)
23. Baptista, R., Antunes, M., Shabayek, AER., Aouada, D., Ottersten, B.: Anticipating Suspicious Actions using a Small Dataset of Action Templates. *13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP)*. (2018)
24. Demisse, G., Aouada, D., Ottersten, B.: Deformation Based 3D Facial Expression Representation. *ACM Transactions on Multimedia Computing, Communications, & Applications*. (2018)
25. Oyedotun, O., Demisse, G., Shabayek, AER., Aouada, D., Ottersten, B.: Facial Expression Recognition via Joint Deep Learning of RGB-Depth Map Latent Representations. *IEEE International Conference on Computer Vision Workshop (ICCVW)*. (2017)
26. Müller, M.: *Dynamic Time Warping. Information Retrieval for Music and Motion*. Springer. (2007)
27. Chu, WS., Zhou, F., De la Torre, F.: Unsupervised temporal commonality discovery. *European Conference on Computer Vision (ECCV)*. (2012)
28. Magee, D.J.: *Orthopedic physical assessment*. Elsevier Health Sciences. (2014)