HOCHSCHULE FÜR TECHNIK UND WIRTSCHAFT DRESDEN
UNIVERSITY OF APPLIED SCIENCES

Proceedings of the 23th Bilateral Student Workshop CTU Prague

Dresden (Germany) 2019

Editors: Hans-Joachim Böhme
Frank Bahrmann
Miroslav Macík
Jan Balata

23th Bilateral Student Workshop CTU Prague

Hochschule für Technik und Wirtschaft Dresden

Contents

Incremental Pose Estimation of multiple LiDAR Scanners using their Pointclouds

Jonas Fabich

Faculty of Mechanical Engineering,
University of Applied Sciences Dresden, Germany

Abstract. To fully utilize a system of multiple lidar scanners a combined point cloud of the sensors is needed. A high quality and robust point cloud registration is going to be achieved in a more difficult environment then existing processes are suitable for. In this work multiple existing registration methods and point cloud feature descriptions are combined in a new way to achieve this.

1 Introduction

For more demanding surveillance or measurement problems multiple LiDAR scanners can be used to eliminate typical single scanner problems. For example the decreasing point cloud densities with increasing distance or information loss in shadows. There is a very exact pose estimation needed to combine the point clouds for further use. To provide an easy, flexible and inexpensive solution the existing sensors can be used to estimate their relative poses using their output point clouds.

2 Related Work

This work builds onto existing registration processes which are mainly suitable for high point cloud densitys and small transformations. [1] A number of functions of the Point Cloud Library are being used to process the point cloud data. [2]

3 Registration Process

Multiple scanners get registered incrementally to achieve a connected transformation. The registration process for two scanners consists of multiple steps to achieve the final result which will be elaborated in the next chapters.

Ground Plane Matching

For the first step, a common flat ground which is observed by every scanner is needed. A Ransac [3] filter is used to recognize the plane and extract the ground plane coefficients. It takes 3 random sample points and checks how many other points are lying in the plane threshold and repeats this step until the main plane is found. This plane will be used as the new x-y plane in every point cloud and therefore solves 3 of the 6 degrees of freedom.

Feature Computation

To further describe each point of the pointcloud features are used which describe the surroundings of a point. The normal estimation [4] computes a hypothetical normal out of points in a specified radius around the point. The Fast Point Feature Histogram [5][6] describes the surrounding points with their surface normals. These features are position and density invariable and can therefore be used to reliably identify correspondences. [7]

Correspondence Rejection

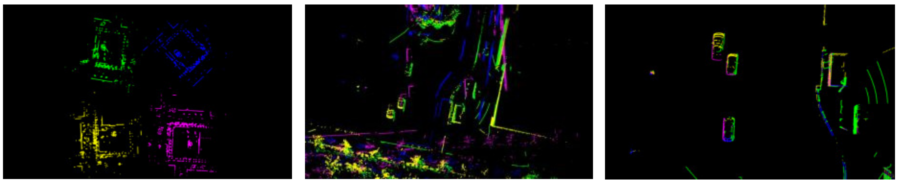
The pcl correspondence rejection [8] uses the point features to support the correspondence search. It also performs a prerejection based on polygonal edge lengths. As a last steps it filters the correspondences by reviewing the resulting transformation. It outputs a rough transformation which is needed for the last step.

Iterative Closest Point Algorithm

The ICP Algorithm [9] is a brute force matching algorithm which can only be used for small transformations. It searches for correspondences, rejects bad correspondences, and estimates a transformation using the good correspondences. These steps get repeated multiple times. The final 3 degrees of freedom are solved by the ICP.

Final Registration

The following figure 1 shows the 2 major steps of the registration process. Picture 1a shows 4 point clouds of their respective LiDAR scanners in different colors. The second picture 1b shows a more detailed view of the transformation achieved by the correspondence rejection. Picture 1c shows the final transformation of the scanners. The examples shown here are taken with scanner distances in the range of 10-20m.



a) Overview before registration

b) Point clouds after correspondence rejection

c) Final transformation

Figure 1: Registration process

4 Conclusion and Future Work

The implementation worked, in the initial environment, providing a dynamic solution as well as fulfilling the accuracy requirements. It was already implemented with further computation to achieve the original automatic measurement objective. The algorithm could be improved to work without the requirement of a common ground plane. It is also still not robust enough to cope with the second test environment, a very sparse scene with larger transformation distances.

References

[1] "Pcl registration api."http://pointclouds.org/documentation/tutorials/registration_api.php#registration-api.

[2] "Point cloud library." http://pointclouds.org/.

[3] R. C. B. Martin A. Fischler. https://doi.org/10.1145/358669.358692, June 1982. Communications of the ACM 24: 381–395.

[4] R.B. Rusu, "PCL normal estimation." http://docs.pointclouds.org/1.7.1/classpcl_1_1_normal_estimation.html#details.

[5] R.B. Rusu, N. Blodow, M. Beetz, "Fast Point Feature Histograms (FPFH) for 3D Registration," May 2009. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Kobe, Japan.

[6] R.B. Rusu, A. Holzbach, N. Blodow, M. Beetz, "ast Geometric Point Labeling using Conditional Random Fields," October 2009. In Proceedings of the 22nd IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), St. Louis, MO, USA.

[7] R.B. Rusu, "PSemantic 3D Object Mapsfor Everyday Manipulation in Human Living Environments." http://mediatum.ub.tum.de/doc/800632/941254.pdf, September 2009.

[8] A. G. Buch, D. Kraft, J.-K. Kämäräinen, H. G. Petersen and N. Krüger, "Pose estimation using local structure-specific shape and appearance context," 2013. International Conference on Robotics and Automation (ICRA).

[9] R.B. Rusu, Michael Dixon,"PCL iterative closest point." http://docs. pointclouds.org/trunk/classpcl_1_1_iterative_closest_point.html#details.

Soft- and Hardware Developments for Immersive Learning

Gunther Göbel and André Minz

Faculty of Mechanical Engineering University of Applied Sciences Dresden, Germany

Abstract. Virtual or augmented reality offers new possibilities for teaching and learning, as students can "immerse" in an artificial environment optimized for a specific task. However, several obstacles exist that currently inhibit academic use of immersive learning, including resource demand and content creation effort. The paper proposes "VR-assisted teamlearning" as a suitable method to reduce resource demand for groups. Also, a soft- und hardware combination is presented that simplifies content creation and interaction with real-world experimental equipment. Two examples are presented that demonstrate the use of these ideas and results as well as applicability are discussed.

1 Introduction

Immersive learning as an effective teaching method is, although already known and analyzed for many years [3] currently getting much more attention [2]. Initiated by the release of consumer-oriented virtual-reality (VR) hardware in 2016, recent hardware for virtual or augmented reality applications is finally creating a reliable and continuously improving basis for immersive learning since then. Immersive learning is currently seen as a solution for mainly three tasks:

– a method to display objects, processes or information in a more intuitive or simplified way,
– a method to display systems that are not physical available, but needed for teaching,
– a method to allow interactions of users in a new way.

The latter is mainly applied for "virtual classroom" applications of users who are physically at different locations, while the other two are aiming at local users or single user experiences. Of course, all tasks can be combined, e.g. a whole group with individuals at different locations interacts on objects via a simplified virtual system for teaching.

2 Challenges of Immersive Learning for Academic Use

Current systems, that address one or more of the above mentioned tasks, are often not well adapted to the needs in university environments for practical courses. Within such lectures, attending students are traditionally split up into smaller groups to jointly conduct experiments, with team sizes around 5-10 persons. Using immersive learning would either require head-mounted-displays (HMDs) for everyone to interact in a single- or shared virtual world, or, with less hardware- effort, require to take turns for a single-user experience. Both solutions are not adequate, as they are either resourcehungry (HMDs and VR-space for all) or time consuming (taking turns). Thus, new ways must be developed to use immersive learning efficiently for teams.

3 Asymmetric "VR-assisted Team learning"

A solution to integrate classical teaching methods better with the immersive ones would be to interface both worlds: if real experiments could have virtual co-workers and all work together on a joined experiment, only few HMDs are needed and no waiting times are necessary. However, the soft- and hardware must allow the necessary interactions: real-world information (like measured signals) must be made available in the virtual world and "virtual" decisions must have an effect on real machinery.

4 Example Application "Welding Trainer"

A comparable approach is already in use: As published earlier [1], the HTW Dresden was one of the first university in Germany to include immersive learning methods for teams into standard university courses. Besides chemical interaction simulation and assembly-training, manual welder training is used in the VR-labs in the mechanical engineering faculty. Using an inhouse-developed software, a VR environment was implemented to avoid safety issues when teaching first steps in welding to students.

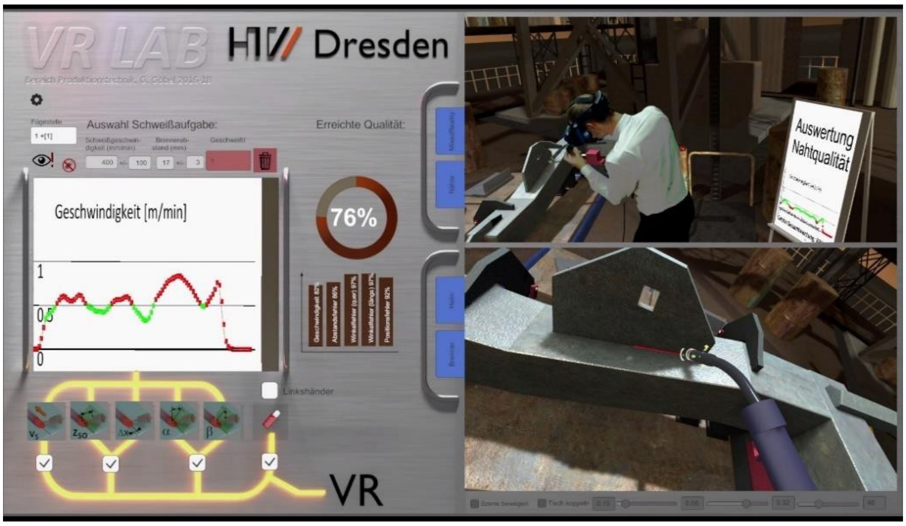


Fig. 1. weld trainer screenshot: shown is the real-time analysis screen for supporters, the current welder-user is visible within the “mixed reality” view in the top-right subimage, his eye view below

This system implements a rudimentary form of asymmetric team learning: one welder (in VR) is supported by 2-3 other students (on screen), see Fig. 1. They not only see the welder view but have access to various additional welding-analysis methods in real time. After a while, roles are switched. It was shown, that already this active involvement from outside the VR scenery helped to reduce beginner errors later when the welder-role was assumed. Students could, due to the running effort of their peer, identify strongly with the task and were motivated to understand and master the process as quickly as possible. As the computer already evaluates real-time results in relation to quality requirements, a much more precise help can be given to the welder than with real welding trials. Important process information (like current moving speed, nozzle distance…) are hard to judge even for experts on real welding, while they are easy to determine and correct within the VR-technology. Thus, the learning curve is steeper even with less trained “trainer”-personnel. Since 2016, the courses are part of the standard curriculum for all mechanical engineering students. The system helps to speed up the teaching process for Gas-Metal-Arc-Welding (GMAW) from theory to real welding significantly. It was concluded that VR-assisted team learning systems are helpful, but implementation needs to be made easier for academic staff to address a wider audience.

5 HTW VR-Framework

The internally developed "VR-Framework" at the HTW Dresden was designed to address the above mentioned needs to simplify further development of VR or AR learning experiences. The following capabilities where determined as necessary for the framework:

- asymmetric team interaction: VR + non-VR tasks can be linked together
- real-world measurements and signals can be transmitted into VR and out of it
- non-VR tasks should allow easy integration of standard teaching material (learning texts, pictures, videos, quizzes…)
- standard learning-content file formats should be importable to reduce conversion efforts
- reduction of computer requirements: non-VR interfaces should be simple and run on any device
- adaptability: if possible, the whole learning environment should be editable by the lecturer, this editing should not require programming skills but should rely on visual programming, drag-and-drop and WYSIWYG editing

Several commercially available systems were surveyed at the beginning of the process but did not offer the required flexibility at the time. Therefore, a new system was developed and the above mentioned requirements were integrated. The basic concept is described in Fig. 2:
As can be seen, the main system uses a backend that drives not only the VR- HMD but also a (WYSIWYG-editable) HTML front-end as a 2D-interface for the supporting team. Hardware-wise a (possibly large) touch-screen is recommended to ease interaction for the non-VR users. A network-connected Raspberry-Pi is integrated to realize in- and outputs with real machinery. Thus, even virtual-reality interfaces can be used to drive real systems and their output can be mirrored back to control 2D- or VR-signal visualizations. The Raspberry-related methods were written in C++ run on a standard Raspbian distribution. General information about the experiment, like status information, helpful background information as well as current signal data are shown on the HTML-front of the main PC. Hints as well as current and completed tasks are all displayed in an integrated quest-system that tracks user progress and initiates new tasks. As intended, the underlying logic of the experiment (e.g. "On click Button A show hint B and load VR-object C") can be edited visually in the browser using so called "blue-prints". While this makes it easy to adapt pre-made experiments.

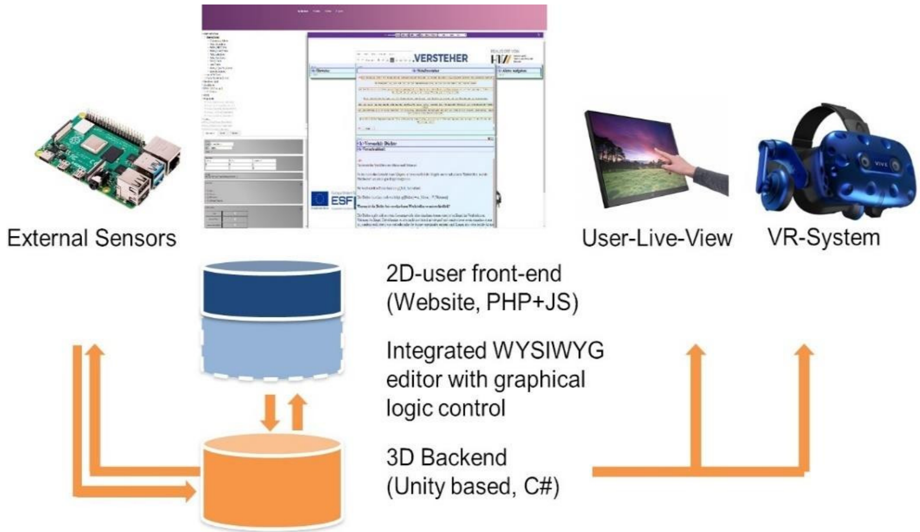


Fig. 2. Fig. 2: overview of the "HTW-VR-Framework" concept that connects an editable VR-World with a 2D-user front-end as well as with external input/output-signals (realized via Raspberry Pi)

Also the implementation of new experiments is facilitated: here a direct import of objects and their components from the Unity 3D development software was foreseen. Their respective blue-print is created automatically and they can take part in the visual programming. A recompilation of the software with the new objects is NOT necessary. Thus, the system is much more flexible to use than typical one-use-only VR-experiences that require full SDK knowledge in Unity, Unreal Engine etc. for any adjustments.

6 VR-Framework Example Application "Metallversteher" ("Metalexpert")

Based on the VR-Framework, several learning experiences are currently evolving for teaching in the HTW. The following example was applied within a funded project and it helped to deduce several general recommendations for VR-assisted team learning. The goal of the experiment is to increase learning effect in the field of metal physics and material know-how by using VR-elements. Teams up to 3-5 persons are able to use the system with just one HMD, see Fig. 3. Tasks are split between real-world experiments (like weighting unknown samples or connecting circuits) and VR-parts (like heating up samples virtually).

To increase team- sense and to immerse also nonVR members into the experiment more, a storyline driven mainly by the VR-world is used. To solve the challenge (in the example a specific metallic objects needs to be identified with the help of experiments), the team must work together to help their member in VR from the outside.

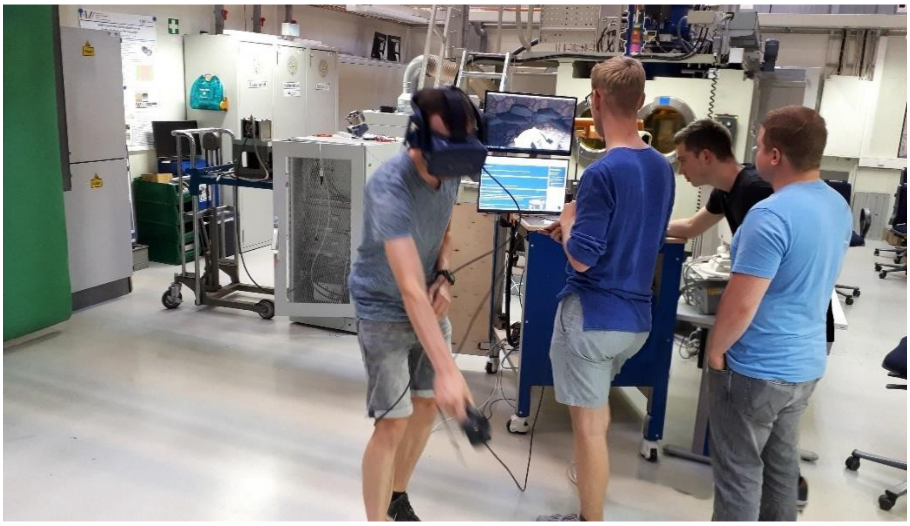


Fig. 3. Mechanical engineering students use a HTW-Framework based application: the 2Dwebsite interfaced with the VR-world allows an easy to create, story-driven team experience incorporating virtual-reality elements

Among others, the following generalized observations were made during the project evaluation:

- The immersion from the VR-scene passes successfully over to non-VR users if they are strongly involved in the outcome of the current scene (e.g. by giving relevant information, solving questions…)
- The balance of story-related VR-scenery and learning-related objects must be carefully kept to avoid unnecessary distraction
- One of the drawbacks of VR, that it is hard to see and understand for spectators what is going on inside can be used to an advantage: it creates curiosity and increases communication. Missing information must be conveyed by the current VR-user, forcing him to reflect more on her or his surroundings
- VR-scenes should be created in a way to help switch roles during the experiment. Thus, every team member can be VR-user for a certain time, which keeps the experiment interesting for everybody
- Audio hints or commentary, although generally recommended for VRenvironments (to deal with the unknown visual focus of the VR-user) are not useful for team-assisted VR. As tasks run parallel in VR and real-world, audio information is often missed in cross-talk
- If possible, support team and VR-use-space should not be too close to reduce unnecessary side-communication and distraction

According to questionnaires filled in by the users, 96% attested a high or very high interest in such experiments and 93% perceived the content as practically applicable. Users agree that especially interactable objects in VR felt real enough to remember them equally well like real world experimental gear. Distraction (by their team mates and the general VR-scenery) was discussed as a problem that lead to missing information. Here a careful scene design is vital.

7 Conclusion

Based on the general idea of immersed learning, specific needs were deduced for its applications in an academic environment for practical courses. The concept of "VRassisted team learning" was introduced to minimize hard- and software effort for maximized learning with student groups. It was tested in several examples. It could be shown that students are highly motived by such content and a directly applicable learning effect is possible. A new helper hard- and software was designed and implemented to reduce efforts to create immersive learning experiences for teams. This so called "HTW-VR-Framework" was already used to examples successfully. The optimal way to make this framework available to a larger user-base is currently in discussion.

Acknowledgement

Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des vom Sächsischen Landtag beschlossenen Haushaltes.

Bibliography

[1] Sonntag R. Göbel G. Experiences and acceptance of immersive learning arrangements in higher education. In *International Conference on ICT Management for Global Competitiveness and Economic Growth in Emerging Economies*, Wroclaw, Poland, 2017. October 23-24, 2017 Proceedings.

[2] Thomas Hutzschenreuter and Christian Burger-Ringer. Impact of virtual, mixed, and augmented reality on industries. 2018.

[3] Merchant Z. Goetz E. T. Cifuentes L. Keeney-Kennicutt W. Davis T. J. "effectiveness of virtual reality-based instruction on students' learning outcomes" in k-12 and higher education: A meta-analysis. *Computers and Education*, pages 29–40, 2014.

Qualitative comparison of methods for example-based style transfer

Michal Kučera

Faculty of Electrical Engineering,
Czech Technical University in Prague, Czech Republic

Abstract. In this paper, which I have done as my master's thesis, I provide insight into the current state-of-the-art of example based style transfer methods, and design and perform a series of tests to compare the differences of output qualities of the studied methods.

1 Introduction

In the last years filmmakers have been trying to introduce visual styles similar to paintings and cartoons into their movies with varying degrees of success. *Loving Vincent (2017)* was the first fully hand-painted movie, where a group of painters created the whole movie frame by frame, but even though the movie was a commercial success, it also showed the infeasibility of creating these kinds of movies due to the extreme time requirements and cost. Since 2014, computer graphics researchers started coming with their takes on algorithms that could make this process easy, quick and cost-efficient, and while the first iterations of these algorithms were too flawed to be used commercially, current state-of-the-art algorithms are now being used by major film studios as well as in some of the more popular video editors (*Adobe After Effects* for example). But during these years, the research done in this area has started to branch out, creating several completely different approaches to this problem which all differ significantly in their requirements and workflow, but also in the quality of their stylized outputs. And while we can easily compare their HW requirements to understand their advantages and disadvantages, the output quality is also very important aspect to consider. Unfortunately, given that the perceived quality is completely subjective, measuring output quality is a difficult task. I have decided to tackle this problem by designing and performing several experiments with users, so that we can have an idea how do the current state-of-the-art methods compare to each other.

2 Related Work

While there are some papers comparing different style-transfer methods side by side, they only focus on one very specific branch of methods (only patch-based, only neural-based etc.) and they are not doing any experiments to compare the quality of their outputs. Lack of this kind of study was the main reason why I have decided to select this as my topic.

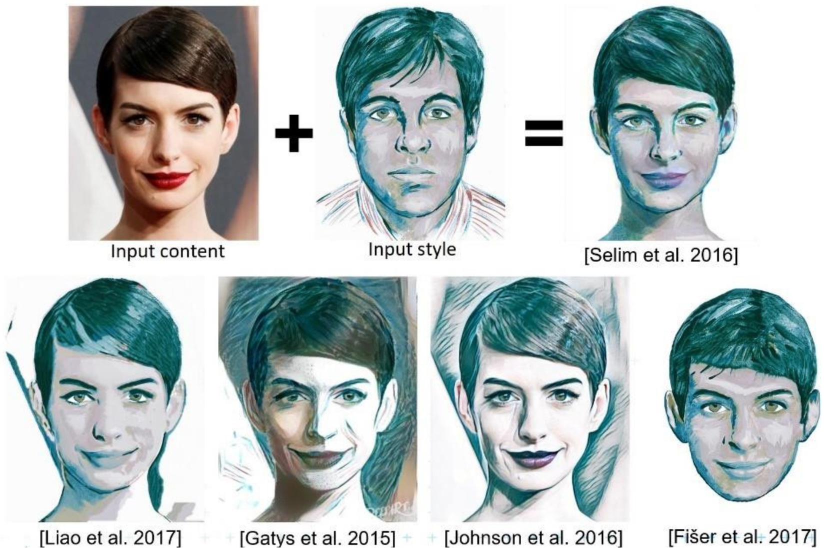


Fig. 1. Outputs of several tested style-transfer methods with the same input content and input style.

3 Perceptual experiments

First thing I had to do was creating a suitable dataset for testing. Unfortunately, authors very rarely use the same combinations of input contents and styles in their papers, as they commonly use those where their result looked the best. That meant that I had to implement each of their methods based on their papers or existing and available implementations, and by trying their own inputs and results verifying the correctness of my implementation. After implementation was done, I have selected input contents and styles by selecting some from each paper used, so that every method had some they perform well on and some that other methods perform well on. The dataset was then created by making outputs for every combination of input content and style.

After creating the dataset I have performed a testing experiment to verify the experiment's design, which was strongly based on basic user tests that are commonly performed when testing design of web pages. By selecting 13 people where each of the groups of interest was represented equally (men/women, artists/non-artists, programmers/non-programmers etc.) and running a moderated in-person test, I have verified the design and then ran the test on much larger scale by creating a survey website and releasing the test to the public.

4 Conclusion

Through the survey around 1500 answers were collected, which were used to evaluate qualitative differences between methods. The experiments strongly suggest that patch-based methods are now better in output quality than neural-based methods, which confirms what was already believed in the industry (commercial video editing SW only use non-neural approaches). Results from this thesis and experiments used can now be used as benchmarks for the output quality of various style-transfer methods and can help with decisions which methods to use in specific cases.

Experiments in this thesis were later used as a part of *Futschik et al. 2019* paper on video stylization.

External Labeling With Utilization of Containment Information

Václav Pavlovec

Faculty of Electrical Engineering,
Czech Technical University in Prague, Czech Republic

Abstract. The presented algorithm is able to position textual annotations in the free space in medium complexity scenes in real time. Contrary to previous approaches, the free space for a given object is extended by the area of its containing object. Therefore, the positioning of the labels is more versatile.

1 Introduction

Complex objects consisting of large number of components appear in multitude of domains. Illustration of these objects conveys information about spatial arrangement of their parts. In order to communicate the function of the depicted parts, textual description is required. Labels represent the means of connecting the textual description with the visual representation of the parts.

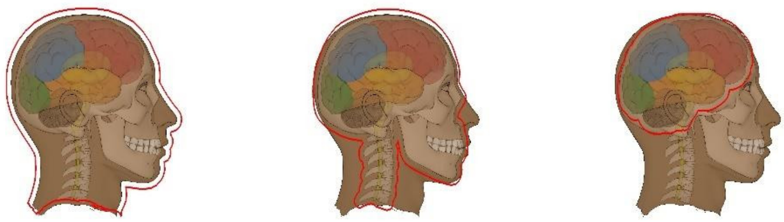
The approach of Čmolík and Bittner [2] positions the labels around a convex silhouette of the object. Its main disadvantage is the length of leader lines, which was addressed by Pavlovec [3], who replaced the convex silhouette by an extruded boundary of the object. However, illustrators utilize an even wider range of positions of external labels, if necessary. They can place the labels almost anywhere they see fit as long as there is no visual interference, and the labeling respects the structure of the object. In this work, I focus on further reduction in leader line length by introducing new areas where the label can be placed. If one part contains another part, then the labels for the first part can be placed over the second.

2 Labeling Algorithm

The following section briefly describes the presented algorithm. The algorithm is similar to that described by Pavlovec in [3]. However, in this approach, I place the main focus on the utilization of containment and its benefits. The algorithm is a screen-space technique and performs heuristic-based optimization aimed to find optimal label positions with respect to a set of criteria. There are several general requirements for external labels that need to be met. According to Ali et al. [1], labeling should meet the following criteria.

1. Readability
2. Unambiguity
3. Compactness
4. Aesthetics

Label Placement which takes into account the structure of the object, i.e., hierarchical information will inevitably produce label layouts with higher compactness. External labels are placed mainly around the labeled object, where there is free space for their placement. Figure 1a shows these potential labeling positions as the outline of the first level of the hierarchy, i.e., the outline of the whole object. However, as Figures 1b and 1c show, there are more levels of the containment hierarchy and, therefore, more possible labeling positions as depicted by the red outlines of the parts contained by another larger part.



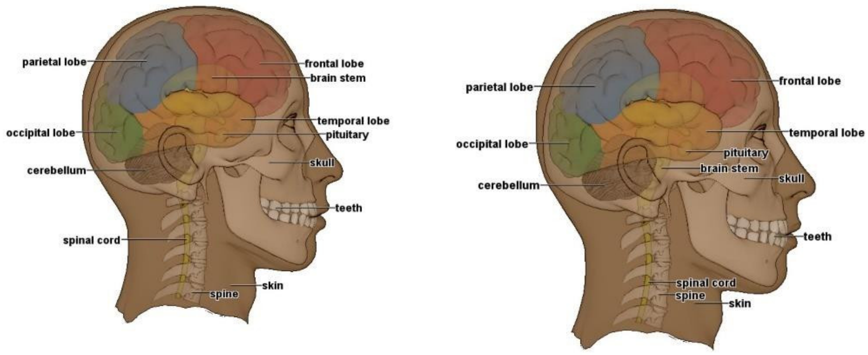
(a) (b) (c)

Fig. 1: Outlines of parts on different levels of containment hierarchy. (a) Level 1, container: screen. (b) Level 2, container: skin. (c) Level 3, container: skull.

Obtaining the possible labeling positions is not a hard task with knowledge of the part hierarchy. The hierarchy can be obtained in multiple ways. It can be a user input or a hierarchy gained from the 3D modeling tool, used to create the model in the first place. Nonetheless, this way of acquiring the containment hierarchy of the parts is not preferable, since the process cannot be automated. I need to deal with non-water-tight objects as well. Therefore my approach is to construct the graph representing the containment hierarchy as a human user would approach this task; By examining the object from different views and forming a wholistic idea from them. For every single view, I calculate how every part occludes every other part and output this information in a graph representation. I do this by line scattering [4] in the process of rendering the ghosted view. Afterward, I aggregate the graphs obtained from individual views into a single aggregated graph. The edge is present in the aggregated graph, is it is present in at least $100*t$ % individual graphs, where $0 \leq t \leq 1$ is a thresholding parameter. The aggregated graph represents the containment hierarchy from a human perspective and yields the labeling sites from Figure 1.

3 Results

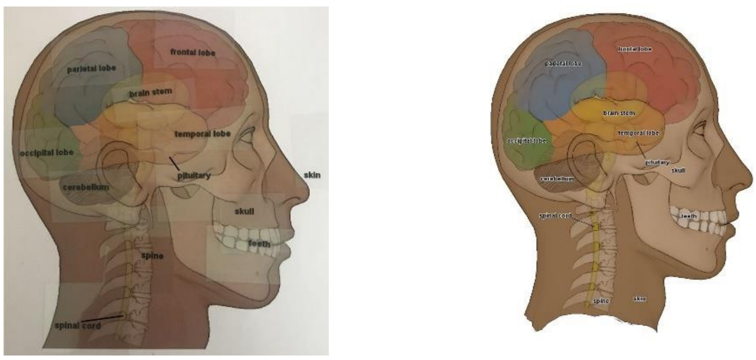
This section presents the results I obtained with the algorithm described in Section 2. Figure 2 depicts label positions with different preferences for the length of the leader lines.



(a) (b)

Fig. 2: External labeling of the head with utilization of containment. (a) Weak preference of compactness. (b) Stronger preference of compactness.

Figure 3 compares the results provided by an illustrator with an output of my algorithm attempting to mimic the label layout of the illustrator.



(a) (b)

Fig. 3: Demonstration of the algorithms capability to match the results generated by a professional illustrator. (a) Labeling created by a professional illustrator. Label positions were of interest, not the illustration itself. (b) Labeling produced by my algorithm attempting to duplicate the results of the illustrator.

4 Conclusion and Future Work

The presented method is capable of positioning external labels for 3D models while respecting the hierarchy of an object. For this purpose, I compute a graph of the containment hierarchy. The graph is computed as an aggregation of perceived hierarchy from multiple distinct views.

In the future, I would like to utilize the containment information to produce ghosted views of 3D objects automatically by modifying the transparency of individual parts.

Acknowledgement

This research has been supported by MEYS of Czechia OP VVV grant No. CZ.02.1.01/0.0/0.0/16 019/0000765 – Research Center for Informatics and by Grant Agency of CTU in Prague grant No. SGS19/178/OHK3/3T/13 (FIS 13139/161/1611937C000).

References

[1] Ali, K., Hartmann, K., Strothotte, T.: Label layout for interactive 3d illustrations (2005)

[2] Čmolík, L., Bittner, J.: Layout-aware optimization for interactive labeling of 3d models. Computers & Graphics 34(4), 378–387 (2010)

[3] Pavlovec, V.: External labeling of 3D objects (2017)

[4] Scheuermann, T., Hensley, J.: Efficient histogram generation using scattering on gpus. SI3D 7, 33–37 (2007)

Real Time Viewing Direction Analysis to Store Recognized Faces

Richard Lehmann, Frank Bahrmann, Falko Lischke, and Hans-Joachim Böhme

Department of Artificial Intelligence / Cognitive Robotics,
University of Applied Sciences, Dresden Deutschland

Abstract. This document describes a simple way to estimate the deflection of a human face under real time condition. The algorithm was evaluated with self-recorded images.

1 Introduction

In the past years service robots became an increasing element of the daily life in different parts of the society. Since they often interact with people, it is pretty logical that the robots shall recognize people by their face. This aim is especially important if the robots are used in sensitive workspace, e.g. in a medical treatment. It is necessary that the patient can build an emotional bond to the robot. This can be achieved by automatically recognize people and call them by their name.

It would be fatal if the robot calls a patient by a wrong name, especially for people with dementia. For this reason the image database should be automatically maintained, in a way that a new image is always saved if the person in the new image looks more straight in the camera than in the image which is already stored within the database. Thus we secure that the robot can use as many facial features as possible to distinguish the persons.

The following work will focus on explaining how a strong face recognition database, that fits the special needs for working with mentally ill persons, can be created.

2 Goal

In this work we will focus on how an estimator can be built that return an evaluation on how straight a person's face appears in a camera, with the image of the person as the input. Other parts of the necessary work, e.g. how to save recognized faces to the associated person, will not be discussed.

Some key features of the needed estimator are:

- Good maintainability
- Small implementation expenses
- Low error susceptibility
- Real-time capability

3 Setting

Currently a robot, that is maintained by the Artificial Intelligence Research Group at the University of Applied Science Dresden, supports caregivers dealing with people that suffers from dementia in a local care facility.
The robot is able to entertain the elderly people by audio outgoings like presenting the latest news, so the caregiver can focus on more important issues. Perspectively the robot shall require the audience to participate in active dialogs. To achieve this the robot have to keep the people apart by their face. There are already several cameras mounted on the robot, that can be used for face recognition and identification. These cameras are a kinect and two 180-cameras, which are interpreted as one omnidirectional camera.

Fig. 1. *August der Smarte* is one of the currently used robots.

4 Core Idea

4.1 Preliminary Note

The goal is to determine if a face in a picture is more frontal looking, than the face in the current reference image for that particular person.
To achieve this, it's important to understand how a face can be turned away. It shows that there are three ways a head can be turned:

- Yaw
- Pitch
- Roll

Figure 2[1] shows these three dimension in a more intuitive way.

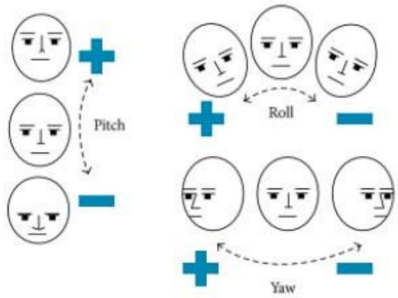


Fig. 2. Illustration of the three angles a head can be turned.

In our case it is quite clear that roll-angle does not have a big impact on the performance of the algorithm, since the main facial features won't get covered. Therefore we want to focus on the two remaining angles, the pitch and the yaw.

4.2 Get Facial Features

From now we know what we like to measure. The next step is to work out how we can use this, to calculate how much a face is turned away. One point is to use the so called *landmark-detection*. This algorithm takes an image and returns a list of x- and y-coordinates of typical points in a human face [2].
One result of the algorithm is shown in Figure 3[2]. Here all of the landmarks are drawn on the original image. Every dot is the location of an estimated landmark. Only the green landmarks are necessary for the viewing direction analysis.

4.3 Calculation

As we have the x- and y-coordinates of every needed landmark we can use them to approximate the yaw- and pitch-angle.

Yaw Estimation To calculate how much a face look to the left or right side, we first calculate the average of x-coordinates of the nose and the chin. We note these with the equation:

$$\emptyset nosechin = \frac{Xchin + Xnose}{2}$$

From here $\emptyset_{nosechin}$ shall be our central x-coordinate. Now we calculate the distance between this central x-coordinate and the x-coordinate of the left eye

[1] Image taken from: https://github.com/jerryhouuu/Face-Yaw-Roll-Pitch-from-Pose-Estimation-using-OpenCV
[2] Original image taken from: https://unsplash.com/photos/wO4gtmdlW5Y

4

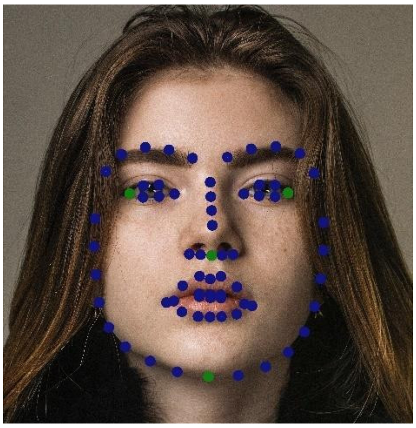


Fig. 3. Every estimated landmark. Only the green landmarks are needed in further calculation.

and the x-coordinate of the right eye. These two distances shall be called $dx1$ and $dx2$. The following formula shows the calculation of these two distances:
$dx1 = \emptyset_{nosechin} - x_{leftEye}$
$dx2 = x_{rightEye} - \emptyset_{nosechin}$
Figure 4[3] shall sketch these idea.

It's easy to see that $dx1$ is smaller than $dx2$, since the person on the image looks to the left side. An image of a person looking to the right side would result in a way that $dx1$ is greater than $dx2$.
We use the relation of these two distances to calculate a new measurement:

$$yawValue = \frac{min(dx1\ , dx2\ )}{max(dx1\ , dx2\ )} \in (0,1] \quad (1)$$

The new *yawValue* is now an indicator of how straight a person looks to the camera in sense of the yaw-angle.

Pitch Estimation To make a statement if the face looks to the ground/ceiling or straight in the camera in sense of the pitch-angle, we use the same ideas as seen before. First the average of the y-coordinates of the left eye and the right eye is determined:

$$\emptyset eyes = \frac{y_{leftEye} + y_{rightEye}}{2}$$

After this we calculate the distance between $\emptyset_{eyes}$ and the y-value of the nose landmark. We note this as $dy1$. We also do this with the distance between the y-coordinates of the nose and the chin, which we call $dy2$. The formulas are the following:

[33] Original image taken from: https://unsplash.com/photos/AxI9niqj_60

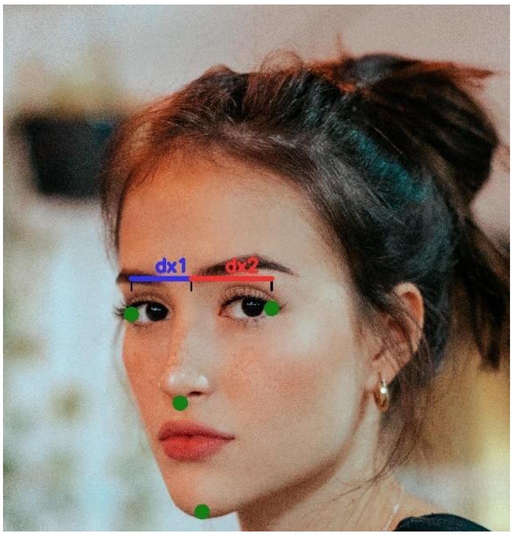


Fig. 4. The green landmarks are used to calculate dx_1 and dx_2, that allows to estimate the yaw-angle.

$dy_1 = y_{nose} - \emptyset_{eyes}$
$dy_2 = y_{chin} - y_{nose}$
These distances are shown in Figure 5[4].

Again we note that as the person looks up, since dy_1 is smaller than dy_2. The relation between dy_1 and dy_2 will be called *pitchValue*, and is calculated in the following way:

$$pitchValue = \frac{min(dy1\ ,dy2\ )}{max(dy1\ ,dy2\ )} \in (0,1) \qquad (2)$$

This value is nothing but an estimation of the pitch-angle magnitude.

4.4 Combine the Results

From (1) and (2) we have the so called *pitchValue* and *yawValue*. In a perfect world where all persons have symmetrical faces, and the landmark estimation would work without mistakes and if the person will look exactly straight to the camera *pitchValue* and *yawValue* would be each 1. We use this idea to define the following equation:

$$Q = \frac{pitchValue + yawValue}{2} \qquad (3)$$

Where Q shall be our quality result. We note the following things:

[4] Original image taken from: https://unsplash.com/photos/i3b2annf4bY

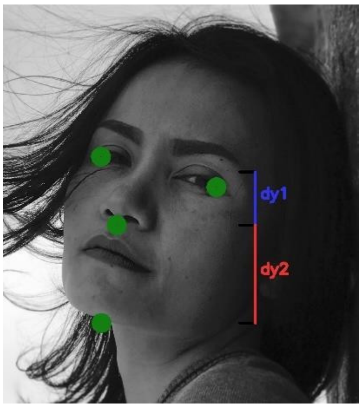


Fig. 5. The green landmarks are used to calculate dy_1 and dy_2, that allows to estimate the pitch-angle.

$Q \rightarrow 1$ . . . the face is quite central
$Q \rightarrow 0$ . . . the face is strongly turned away
With Q we now have a value between 0 and 1 that helps us to estimate how much a face is turned away.
After recognition of a known person by their face, the current image and the old reference image each get a Q-value: $Q_{current}$ and $Q_{database}$.
If $Q_{current}$ is greater than $Q_{database}$, the old image is replaced by the current one.

5 Implementation

The implementation itself was done with python. This decision was made, because there are a lot of useful libraries for the python language.
One used library was face recognition[5], which brings an interface for the facial *landmark-estimation*. The used implementation based on the Paper described in [1], and was trained with the dataset [3].
The implementation of the algorithm, presented in section 4, was quite simple. Hence this the targeted key features noted in section 2 are absolute reachable.

6 Verification

In the previous chapters we introduce a simple way to estimate the deflection of a human face. In this part we like to evaluate this process.
For this reason we carry out a small experiment. First we organize a sensor, a so called gyroscope, that is able to measure the angular velocities in all three

[5] https://github.com/ageitgey/face_recognition

dimensions: pitch, yaw and roll.
We used the CC2541 sensortag from Texas Instruments, seen in Figure 6[6].

Fig. 6. The CC2541 sensor was used for verification.

We fix these sensor on one persons head. This person can move his head, while the gyroscope sends the current angular velocities to a computer. During this, a webcam takes pictures of the moving head. In this way we can achieve a one-to-one relation between the image and the angular velocities at the time the photo was taken.
After this process, we can plot the three dimensions (pitch, yaw, roll) for every taken picture in a three dimensional coordinate system. We choose a dot as marker for every image in this coordinate system. The color of these dots are defined by the Q-value calculated by (3).
The diagram in Figure 7 shows the mentioned coordinate system.
If a dot has a color near red, it means that this point is labeled as a picture that

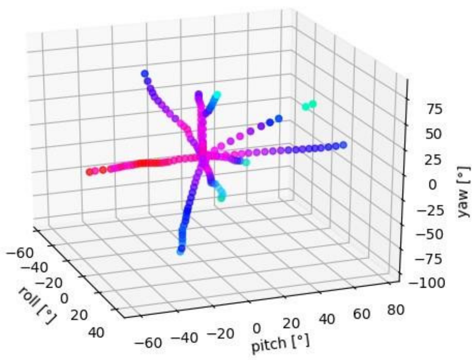


Fig. 7. Q → 1 results in a more red color. Q → 0 results in a more blue color.

[6] Image taken from: http://www.ti.com/tool/CC2541DK-SENSOR#2

contains a face that looks frontal in sense of 4.1. As the color gets more to blue the associated picture of the face is labeled as not looking straight to the camera.
We see that in the coordinate center, where the three angles are near zero, the colors of the dots are violet. Their color goes more to red than to blue.

To gain a further overview we look at the two dimensional representation of this plot without the roll angle, seen in Figure 8.

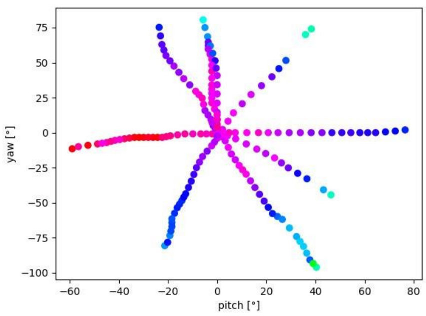


Fig. 8. The colored datapoints, that represents the images, in a two dimensional coordinate system.

It shows that in the negative pitch direction the viewing direction analysis quality gets better labeled. This is fairly unusual, especially if we look at the other movement axis, which all are quite as expected.
This behaviour was confirmed in at least ten tests. Possible error sources might be:

- Measurement errors in the sensor
- False position of the camera

With this small experiment we have a evidence that, under constraints, the calculation presented in 4.4 is applicable.

7 Discussion and Outlook

As we see in Figure 5 at the chin landmark, the used *landmark-estimation* algorithm works not exactly. Since we just like to estimate how much a face is turned away, and not in which direction, these inaccuracies are acceptable. For a further work this problem could be avoided by using a more precise implementation. Another way would be taking more landmarks into account during the calculation.
If the requirements change in upcoming situations, it is possible that some more

advanced techniques like deep neural networks or some regression algorithms are needed.
But since the results of the currently used algorithm is satisfactory, there is no need to use more powerful methods.

References

[1] Kazemi, V., Sullivan, J.: One millisecond face alignment with an ensemble of regression tree (2014)

[2] Kopp, P., Bradley, D., Beeler, T., Gross, M.: Analysis and improvement of facial landmark detection (03 2019)

[3] Sagonasa, C., Antonakosa, E., Tzimiropoulosb, G., Zafeirioua, S., Pantic, M.:300 faces in-the-wild challenge: Database and results. (2016)

Raising living standards of older adults - User research

Tereza Preislerová, Martina Klimešová, Pavla Křivanová and Anastasia Surikova

Faculty of Electrical Engineering,
Czech Technical University in Prague, Czech Republic

Abstract. As health services constantly improve, the number of older adults grows. With the help of technologies and modern medicine these people live longer than their parents. However, older adults still meet various kinds of problems. The main goal for us is to come up with a solution that will raise the living standards of older adults. First part of our work is described in this document and is focused on user research. Research findings will help to determine exactly what kind of problems older adults meet and to figure out if there is a chance and need for us to help them.

1 Introduction

We conducted a research in the Czech Republic where every 5th person of the population is aged 65 or older. According to the Czech Statistical Office older adults formed 19,2% of the population in 2018. This document describes target group on which research was conducted as well as research goals, used methodology and research findings. Second part of our work will be focused on concepts of solutions to discovered problems and will be built on research findings described in this document.

2 Target group

Our target group consists of older adults at the age of at least 65. Our potential target subject is a person who lives alone or with his or her partner at a similar age. We excluded older adults living with their descendants who could help them with their daily routine. We also did not want to include older adults who work full-time, we suppose that they are more self-sufficient than not working elderly person.

Unfortunately, we didn't have capacity to focus on really specific cases. Therefore we excluded all elderly who have serious physical illness such as blindness, deafness or movement inability.

3 Research methodology

We conducted explorative research and spoke to 25 older adults and their caretakers. Among them were people from the retirement home, their nurses, students of the University of the Third Age and their teachers. We also interviewed older adults whom we personally knew and their friends together with caretakers of the older adults with a lot of experience with our target group

Every interview took about 60 minutes and was focused mainly on the following 5 topics.

1. **The daily routine of older adults**
   - How well do they sleep?
   - How do they spend their free time?
   - What does their daily routine look like?

2. **Handling of unexpected situations**
   - What if their TV breaks?
   - What if someone knocks on their door?
   - What if they fall? Are they able to stand back up? Are they able to call for help?

3. **The ability to take care of household**
   - How often do they go shopping?
   - Do they cook daily?
   - How often do they clean the house?
   - Are some chores too difficult for them?

4. **Joyful or stressful situations of the elderly**
   - What hobbies do they have?
   - What did they use to love to do but no longer can?
   - How often do they see their family and friends?

5. **Approach to technologies**
   - Do they welcome new technology?
   - Who brings new technology to their homes?
   - Do they use smartphones or computers? What for? Do they even need to use them?

4 Found Problems

Despite being chosen by the same criteria, every participant was unique and had his own kind of problems he faced in everyday life. However, if we take a step back and look deeper to the initial cause of these problems, we will see that some minor problems of different people can be merged into one major issue for the whole target group. For example, if one older lady goes for small purchase every time she goes shopping and the other one goes only if some of her relatives goes along by car, we can figure out that the reason behind both behaviors is the trouble of carrying heavy purchases. This problem is now common for both old ladies and we can think of a solution that will help them both. By this strategy of looking deeper into what participants were describing, we created a list of 10 most relevant problems of older people.

1. **The trouble with carrying heavy purchases**

   Older adults often do not have a driving licence or a car for their personal use. That means they need to carry the purchase by themselves or rely on help of their relatives. Together with limited mobility, it can mean a serious problem.

2. **Poor sight and hearing**

   The sight and hearing are usually getting worse with the growing age. We need to consider this fact in the case of designing some solutions or user interfaces for older adults.

3. **Fear of falling**

   Older adults do not heal so easily as they used to and they are very well aware of this fact. A fall on the street can be a disaster in the form of a long-time stay in the hospital. Some of them fear so much that they do not want to go outside their room at all.

4. **No interest in getting help**

   Older adults want to feel as self-sufficient as possible, they are very happy if they can help someone too.

5. **Feeling too old**

   Despite being old, nobody wants to feel like it. Many of our participants were denying using equipment for elderly that could help them for example with mobility or other health problems. It made them feel too old and they did not want anybody to see them like this.

6. **Feeling like a burden for their close relatives**

   The worst and the most common fear among older adults is to be a burden to other people, especially to their family. They would rather suffer or try to handle things on their own than to ask somebody for help.

7. **Feeling isolated and alone**

   Some participants felt like they lack social contact with their family or friends. Sometimes the reason was in long distance and mobility problems, sometimes because their loved one were too busy to call or visit. In this case using of new technologies can be crucial for solving this problem, but often older adults had troubles with it as well (see problem #10 in this list).

8. **Negative attitude**

   "This is not for me", "This cannot help me", "I'm too old nobody wants to spend time with me" are quite common phrases among older adults that were either afraid of new things or were isolated from their family.

9. **Feeling hopeless**

   Older people with negative attitude towards life seemed like they lost hope of feeling good in their old age and sounded like nothing will ever change it.

10. **The trouble with new technologies**

    We discovered 2 types of approach to new technologies among older adults. Some of them loved to learn new things but were devastated by small buttons or confused by the way the technology should be controlled. On the contrary some of them denied to use any type of new technologies either according to personal convictions or because of the fear of failure. This group of older adults was more likely to feel too old, isolated and alone and had negative approach to other things as well. In other words, they had most of the problems in this list.

5 Conclusion and Future Work

Despite the fact that 25 participants may seem like a low number, our research was focused on the quality, not the quantity. We were interested in gaining as much information as possible and explore the problems, not to talk to as many people as possible. During the interviews, we spoke with happy and socially active people who really enjoyed life. On the other side, we got the chance to speak with very unhappy and lonely people who had no joy of being retired.

Our main findings speak for itself, we believe that social contact is the main essence of happiness for these people. They want to feel needed, self-sufficient and as a part of the world. It should be our main interest to provide it and come up with new ideas and possible solutions, which is the topic of the second part of our work.

Raising living standards of older adults - Concept

Tereza Preislerová, Martina Klimešová, Pavla Křivanová and Anastasia Surikova

Faculty of Electrical Engineering,
Czech Technical University in Prague, Czech Republic

Abstract. There are many problems older adults from the Czech Republic have to deal with. Our research showed for example problems like fear of falling, loneliness or trouble carrying heavy purchases. The last two mentioned were the main problems our work focused on. We created two concepts that are designed to help to solve these issues. The first one introduces a modern shopping bag with wheels that helps older adults to carry heavy purchases. The second concept's purpose is to decrease the isolation and loneliness of older adults. It introduces an application for remote communication using a TV.

1 Introduction

We aim to raise the living standards of older adults. Make them happier and self-sufficient for as long as possible. The target group of our research and development is older adults. Older adults are 65 years old or older, do not work full-time and live alone or with a partner of a similar age. They struggle daily with a great number of issues. We introduced some of these issues in *Raising living standards of older adults - user research*. This work presents concepts of two possible solutions for selected issues. We interviewed 25 older adults from the Czech Republic. This work presents concepts of two possible solutions for selected issues based on the research.

2 Problems to solve

As mentioned in our previous work: *Raising living standards of older adults - user research*, we found many problems our participants struggle with.
These problems were the most frequent ones:

– the trouble with carrying heavy purchases,
– poor sight and hearing,
– fear of falling,
– no interest in getting help,
– feeling too old,
– feeling like a burden for their close relatives,
– feeling isolated and alone,

– negative attitude,
– feeling hopeless,
– the trouble with new technologies.

Now we present two concepts that propose possible solutions to some of these issues.

3 First Concept: Modern shopping bag with wheels

The first presented concept is the modern shopping bag with wheels. It tries to solve the following issues.

During the research, we talked with 25 people and found many problems they have. The ones we try to solve with this concept are:

– older adults have a problem with carrying a heavy purchase (18 of 25),
– older adults do not want to ask for help (10 of 25),
– older adults do not want to feel and look too old (11 of 25),
– older adults feel like a burden (9 of 25).

Some also mentioned having trouble to bend down to put purchase into the bag.

Many older adults refuse to use any currently available shopping bag with wheels. Even though it would make the shopping easy. They find the bag ugly and embarrassing to be seen in public with. Older adults want nice things and the opinion of others is important to them. The bag also needs to be light to allow easy manipulation even with heavy purchases.

The presented concept consists of the modern shopping bag with wheels, inner bag and matching purse. All three parts have matching modern design to be attractive to the target group. Further research is needed to determine what design is found attractive by older adults.

The purse has a strap so it can be easily attached to the shopping bag with wheels. The inner bag can be taken out, filled with groceries and put back inside. Therefore the user does not have to bend down. The inner bag can be replaced with any bag the user wants.

Many issues (such as carrying a heavy purchase or bending down) could be easily solved by grocery delivery. However, none of the participants of the research used it. Further research is needed to determine why grocery delivery is not used.

4 Second Concept: Application for remote communication using TV

The most important finding, and at the same time the most critical problem that needs to be solved, is the fact that most of our participants feel isolated or lonely. They do not have as many opportunities to see their loved ones as they would like to have. Some of them compensate for that with frequent phone calls, but due to a poor hearing that a lot of our participants suffer from, it is not an option for everyone. We were trying to approach this problem so that the solution will help these people as well.

Our research did not find only negative findings, there was also one positive that all participants had in common - they all loved watching TV. They watch it daily and have no or only minor trouble to control it. Our second concept takes advantage of these findings and introduces an application for remote communication that runs on TV.

There is no need to buy or learn how to use any new technology. Older adults already have all they need at home and they are familiar with it. This helps to eliminate trouble they could have with accepting new technology.

The application uses an internet connection to enable remote communication. The user interface is adjusted for older adults - it consists of big elements, it is simple and easy to understand, the volume is adjustable, etc.

The application cannot replace real human interaction, but it is a very good option for people with mobility problems that spend most of their day at home. The application could bring a completely new way of communication to these people. The application enables its users to:

– have a contact list saved in the application,
– make a call to people in the contact list,
– accept a call from people in the contact list,
– show notifications,
– watch TV shows together while having a phone call,
– accept photos or video sent by people in the contact list.

The function of shared TV watching is one of the biggest advantages of this application. It could be potentially used not only by older adults but also by other age groups.

Another advantage of this solution is that spending time together with older adults becomes easier for their close ones. Many older adults do not hear an incoming call when they watch their favorite TV show. With this application, the notification about the call would appear directly on their TV screen and it would increase the possibility that the user will notice the phone call.

Part of this concept is also a new special remote control. It enables users to control the TV and our new application. It is designed especially for older adults - big buttons, minimalistic design, clear labels, easy-to-understand structure.

5 Conclusion and Future Work

Just in the Czech Republic alone, there are 2 million and in Germany more than 17 million older adults we can help.

For future development, we decided to focus on the application for remote communication using a TV. We plan to create a prototype, test it on the target group and iterate upon it.

The final product is not exclusive to older adults and can be interesting for others as well.

Towards the RoNiSCo Mobile Application

Jan Brose, Frank Bahrmann and Hans-Joachim Böhme

University of Applied Sciences Dresden, Germany

Abstract. Service robotics in nursing facilities need the opportunity to communicate with the staff. To solve this problem for Robotic Night Shift Companion (RoNiSCo) for nursing facilities we use a mobile application. We propose to use a simple designed application on smartphones for the Staff to communicate with the robot.

1 Introduction

Currently in Germany there is a shortage of the nursing staff. This causes many problems especially at night shifts there are less nurses than floors in the facilities. This is problematic because a big part of the residents in these facilities suffer from cognitive impairments like dementia which results in a high workload for the staff. But there is also a risk for the residents that dangerous situations might be detected too late. Many approaches have been introduced to help the staff with their work and assist them.
The work group of artificial intelligence at HTW Dresden is developing an assistant robot to assist the staff with their tasks. This project is realised in cooperation with a local nursing facility in Dresden. One part of the project is to develop a Robotic Night Shift Companion. In this application the robot should be able to patrol floors autonomously where currently no staff can be present. The idea of communication with the staff is that the staff has a smartphone with a mobile application which allows them to get information from and give orders to the robot. The design of this application will be the subject of the following sections.

2 Mobile Application

The design for the application is based on the design of its previous version described in [1]. A new implementation of the old one was necessary because we needed to change the library for the connection to the robot. There were also issues with the Android version, some features of the old implementation are deprecated. We also added a new call the robot function which is based on [2] and should allow to call the robot to a chosen position.

2.1 Requirements

The mobile application must be able to build a connection to the robot. It needs to be able to show status information of the robot. The staff should be able to use the application with one hand, there shall be an option to change the hand of use. Accidental closing of the application shouldn't be possible. If the robot finds a resident in need for help he must be able to inform the staff. A simple solution to call the robot to a position where he is needed is necessary.

2.2 Connection to the Robot

To build a connection to the robot we use a ssh-tunnel. In this connection the robot acts as server and the mobile application as client. There might be multiple clients. If there is a new client it needs to be registered at the server. The server and clients share a heartbeat where the robot sends his status information and the clients confirm they got the message. If a client doesn't answer for a certain time all other clients will be informed that this client went missing. The robot sends an alarm message to all the clients in a situation occurs when help of the staff is needed. The clients will be able to call the robot to a certain position. There will be a login interface in order to connect to the robot which can be seen on the left side of figure 1.

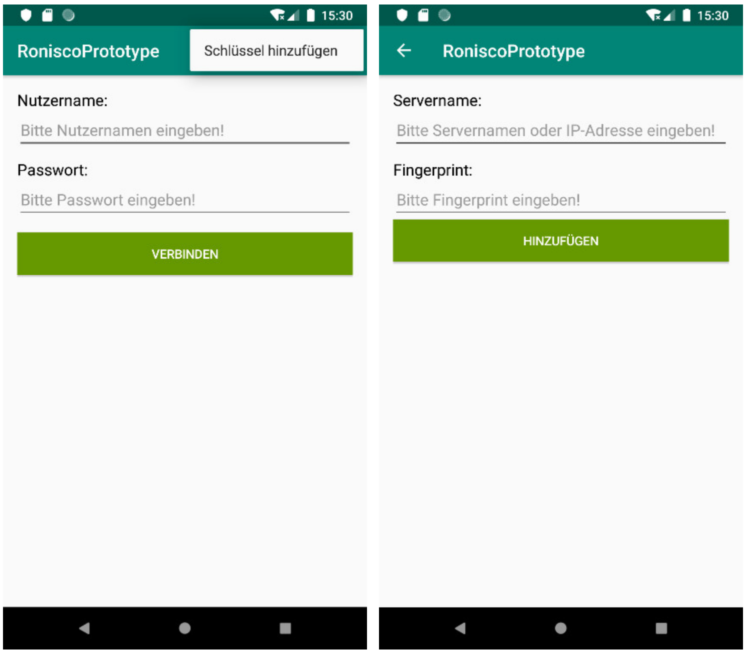


Fig. 1. Left: Login Activity, right: Add Key Activity

If this is the first connection to a new server or the server changed its fingerprint there has to be the option to add a new fingerprint. This is necessary to ensure that the connection is with the robot and not something pretending to be so. This option is given within the menu on the top right corner of the screen. By choosing this option a new dialog appears to add and store a key in the system. This is shown in figure 1 at the right side. Afterwards building the connection should be possible and the user will proceed to the Main Screen. The implementation of an automatic login at the start is also planned because the staff might not want to type the login and password.

2.3 Main Screen

This screen is the center of the application. All other parts are available from here. The screen, which can be seen in figure 2 on the right side, shows some basic information. These are, its current battery charge, a textual description of his location, his current time and a description of his current activities. There is also a button for switching between the two basic modes the robot has (patrolling and ready-state). At the bottom there is a large red call button to open the "call a robot" screen described in section 2.4.
On the top right of the activity there is an option to open a Menu by tapping on the three dots. This Menu can be seen in figure 2 in the central picture, where we can also see two options settings and close connection. Hitting close connection leads to a dialog asking if the user is sure about, beeing seen on the right side of figure 2.

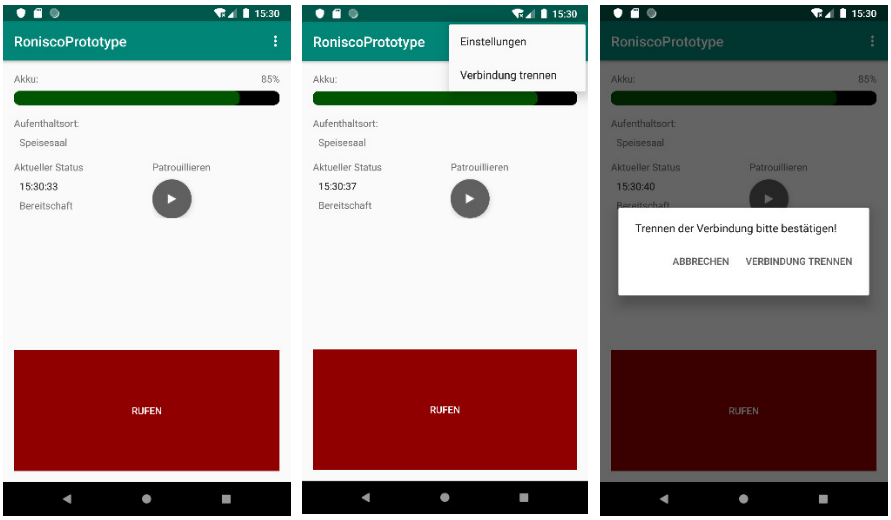


Fig. 2. Left: Main Activity, center: Main Activity with Menu, right: Dialog for Logout

By choosing settings a new screen with the settings is opened. This simply offers the opportunity to change preferred hand for the one-handed use of the application and the zoom-factor of the call a robot part of the application, there is also a button to save the changes

2.4 Call a Robot

This part was necessary because we didn't want that the application gives an opportunity to spy on the staff, we also didn't want to use additional hardware. Therefore simply locating the phone wasn't an option.
An alternativ was given by [2] and it just needed to be adjusted and implemented in android for our work. The idea is to give an exact place by choosing the floor and the room number, which can be entered by a drop-down menu. There is also a Map to determine an exact location and a call button to call the robot to the chosen position. These basic elements were already included in the given prototype.

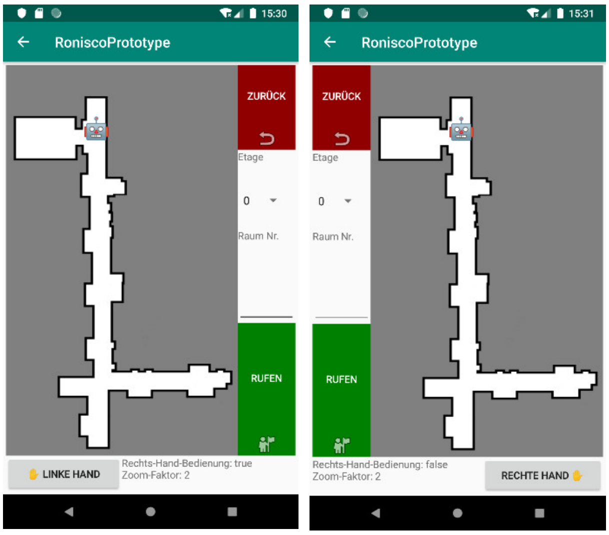


Fig. 3. Left: Call a robot with right hand use, right: Call a robot with left hand use

We extended this layout by adding an option to switch the hand of use below to help in situations where the staff can't use their preferred hand. By pressing the button the layout will adjust to the new hand of use. Therefore the layout will simply be mirrored in the middle of the screen. Both layouts can be seen in figure 3. The map shows the floorplan of the facility. To zoom on the map the user has to draw a figure on the map and the map will be zoomed to the center of the drawing like it is shown figure 4 on the left. The zoomed map also contains some Unicode symbols which shall make it easier to orientate within the plan. There is also the option to add room numbers with residents names, if it is wished, as it can be seen on the right of figure 4.

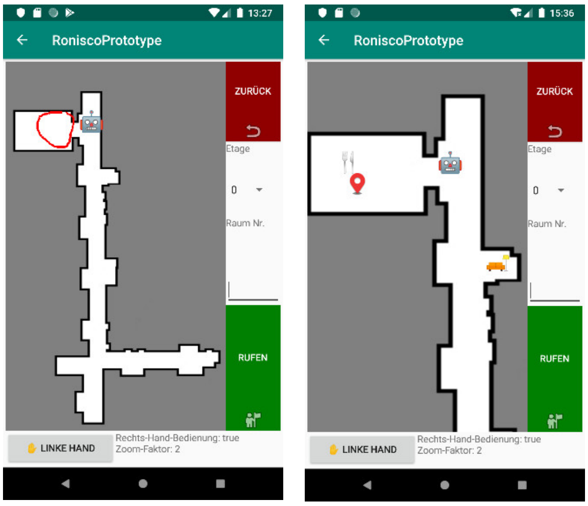


Fig. 4. Left: Draw a figure to zoom, right: Zoomed view

3 Outlook

There are still some points which will need further work. At first there shall be a system to store the unicode symbols for orientation and their location in a more efficient and dynamic way. Then the infrastructure for the ssh-connection needs to be implemented. There is also the need of doing real-world experiments and testing the application.

References

[1] F. Lischke et al. "RoNiSCo: Robotic Night Shift Companion." In: Machine Learning Reports 03 (2017), pp. 26_35.

[2] Dominika Palivcova. "How to Call a Robot in an Emergency Situation." In: *Proceedings of the 22th Bilateral Workshop CTU Prague*. https://nbn-resolving.org/urn:nbn:de:bsz:520-qucosa2-361124, 2018, pp. 37_38.

Development of a Fallen People Detector

Jan Brose, Robert Erzgräber, Frank Bahrmann and Hans-Joachim Böhme

University of Applied Sciences Dresden, Germany

Abstract. To support the staff in nursing facilities at night shifts a robot needs to be able to detect and evaluate situations by himself. Therefore it needs to be able to detect fallen persons in scenes. This students project is based om the RoNiSCo project of the work group of artificial intelligence and cognitive robotic at HTW Dresden. At first we have a look at the current technologies used to solve such problems, afterwards we work out an approach to solve this problem for ourselves. Then we implement some chosen algorithms and compare them. In the following we have a snapshot of the current state.

1 Introduction

Currently there is a shortage of staff in nursing facilities. At the same time many residents of nursing facilities suffer from cognitive impairments. This leads to a high workload for the staff which increases even more at night. It results in a risk that dangerous situations or situations in which residents need help are detected far too late. The work group for artificial intelligence and cognitive robotics of HTW Dresden has a project in cooperation with a local nursing facility to develop a robotic assistant for the nursing staff [5].

One part of the project is a robotic night shift companion application were the robot will be able to patrol floors autonomously, detect and evaluate the situation by himself and call for help if needed. To be able to detect and evaluate these situations besides others the robot must be able to detect fallen persons. In conclusion this project is a two semester students research project with the target of finding an algorithm and concept for a fallen person detection.

The task of the project is divided in two parts. In the first part there will be a literature research towards the topic “fallen person detection“. Afterwards there shall be concepted an approach to implement a prototype of a fallen person detection. In the second part some chosen algorithms will be implemented prototypically and evaluated.

In the following sections there will be a look on the current state of the work at the beginning of the second semester.

2 Literature Research

At the current state two different categories of research emerged. There were papers also using mobile robots, as we do, and the ones which only have basic approaches and used static systems.
The papers where a mobile robot was used came mainly from two working groups. Both use mobile robots with similar dimensions to make and therefore have a better _t towards our scenario than the others.
One is based at the University of Padova in Italy [1, 2]. They used a mobile robot of comparable dimensions as ours. Also their camera module was at a comparable height and the provided data was comparable to ours. Another useful part of their paper is that they showed the advantages of mobile robots towards static systems, which is the fact that a robot has the opportunity to change its own position to evaluate an uncertain situation from another perspective. The robot also has the possibility to use the sensors he uses for navigation to validate the result of their classification of the camera images. To make an easier comparison between new approaches an theirs, they offer a dataset for training and testing. With their approach they reached a $F_{0.5}$-score of 0.86.
The other research group is from TU Ilmenau in Germany [4, 6, 7]. They use a robot of the same manufacturer as we do. But their robot shall operate in a domestic environment. They want to detect if the resident has fallen or otherwise needs help. Their fallen person detection reached a F-score of 0.91. Some similarities have been found especially in preprocessing of the depth images for the classification. In figure 1 can be seen both start with a depth image. Then they have a background subtraction, remove the ground plane and points above a certain height. After that they remove points with none or few neighbors and divide the remaining pointcloud into objects. Afterwards they classify their objects and try to validate their result.
The remaining papers ([3, 8]) can also be divided into two subcategories. Some of them only have basic approaches and no complete algorithms for fallen person detection. The others are to the topic of fall from bed detection.
The problem with the detection of falling from bed like described in [8] is compared to our problem. They know if there shall be a person in the scene. Also they have a limited area where to search for the person.
Only basic approaches were offered by DensePose described in [3] and OpenNI SDK, Microsoft Kinect SDK which were also topic in [8]. Both don't deliver any information towards the detection of fallen person suitable for our scenario and might only be used for preprocessing. With DensePose only using RGB-Data OpenNI SDK and Microsoft Kinect SDK have problems with delivering reliable results if there are unnatural body positions and partly covered persons which also rules them out to use.

3 Approach

To concentrate on the classification itself and not on the background subtraction we used a simpler approach to create data for the classification, as it can be seen in 2. The approach we went for is that we take two images of a scene. At first we have a picture of the scene with a suspicious object and one without. Then we build a difference picture and filter interferences. After this we should have data to use for the classifiers. The data will then be used at the prototypical implementation and testing of the classifiers.

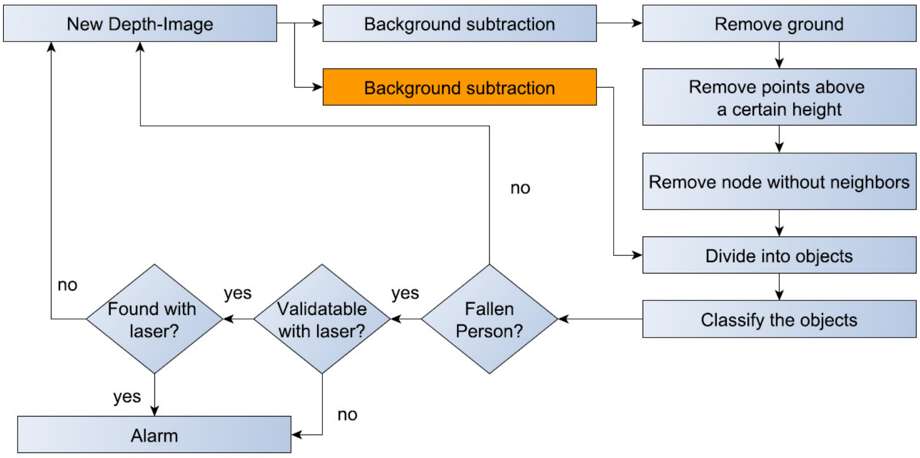


Fig. 1. Blue: basic procedure both groups use, orange: adjustments made to simplify the background subtraction

4 Background subtraction

As previously explained we needed a workaround for the background subtraction. In the next step we build a boolean gridmap of the same dimensions as the depth images. The values of the cells are set to false if the differences between the images are below a certain distance and true if they are larger. Afterwards we erode the image multiple times to erase small interferences. Then we dilated the remaining areas again to expend the remaining areas to their previous size. After this we use a simple floodfill to erase areas below a certain size. In the following we use the gridmap to build a new image. For the cells with a false value in the gridmap we set the pixel to a predefined, negative dummy value. The other pixels get the values from the image with the object. This workflow can be seen in an example scene in figure 2.

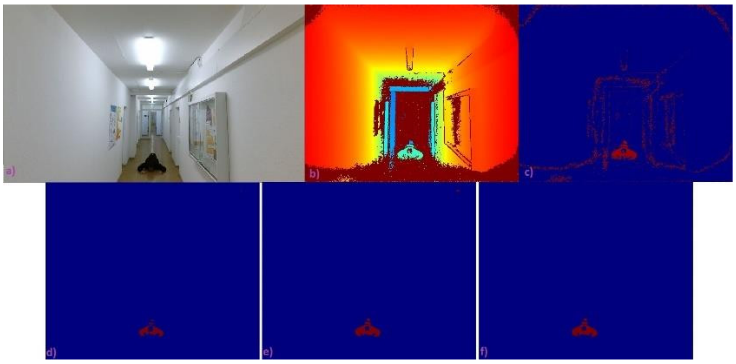


Fig. 2. a) RGB image of the scene, b) Depth image of the scene, c) Difference image, d) Eroded image, e) Dilated image, f) Final image after floodfill

After these steps we have an image we can build a point cloud from and use for the next steps. We also might use these images to locate the object within the other images.

5 Classifiers

For the classification both groups, mentioned in the literature research (section 2), apply different approaches. Both have a point cloud after their background subtraction. In [2] they have a segmentation to Patches of Voxel using Voxel Cloud Connectivity Segmentation. The resulting patches will be classified using support vector machines, like in figure 3 on the left. Afterwards they use euclidean clustering on the positive patches, like it can be seen in figure 3 in the middle. As a next step they use again support vector machines to classify the resulting cluster, like on the right of figure 3.

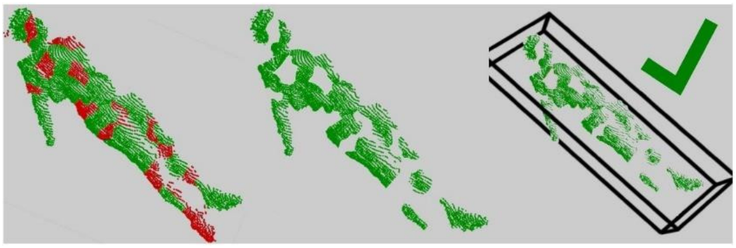


Fig. 3. Source:[2] last 3 steps of the object classification

The result of this classification will then be verified using a map verification if possible.

In [7] they use a RANSAC-algorithm and euclidean clustering for their point cloud. Then they do some transformation and layering where they find the main axis of the object and divide it along this in layers of a certain hight, like it can be seen in figure 4. The layers then will be classified by an Adaboost with Decision Trees as weak learner. If at least two layers were classified as positive they say there is a fallen person.

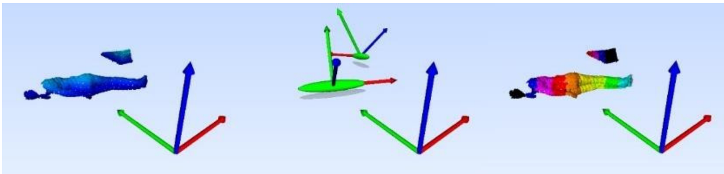


Fig. 4. Source:[7] get the main axis of an object and divide it into layers

6 Outlook

The capturing of more scenes will be necessary in the next steps because we currently only have three of them. After this we need to do the segmentation of the point cloud and then implement the classifiers. For the classifiers we decided to implement the one from [7] at first, because their documentation of how to get the features is better than in [2]. If then there is enough time left in the students project, there will also be an implementation for the approach in [2]. Afterwards the classifiers need to be tested.

References

[1] M. Antonello et al. "Fast and robust detection of fallen people from a mobile Robot". In: *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* Sept. 2017, pp. 4159_4166.

[2] Marco Carraro. "Real-time RGB-Depth perception of humans for robots and camera networks". PhD thesis. University of Padova, 2018.

[3] Rıza Alp Güler, Natalia Neverova, and Iasonas Kokkinos. "Densepose: Dense human pose estimation in the wild". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2018, pp. 7297_7306.

[4] B. Lewandowski et al. "I see you lying on the ground - Can I help you? Fast fallen person detection in 3D with a mobile robot". In: *2017 26th IEEE International Symposium on Robot and Human Interactive Communication* (RO-MAN). Aug. 2017, pp. 74_80.

[5] F. Lischke et al. "RoNiSCo: Robotic Night Shift Companion." In: *Machine Learning Reports* 03 (2017), pp. 26_35

6

[6] Markus D Solbach and John K Tsotsos. “Vision-based fallen person detection for the elderly”. In: *Proceedings of the IEEE International Conference on Computer Vision.* 2017, pp. 1433_1442.
[7] Tim Wengefeld, Benjamin Lewandowski, and Horst-Michael Groÿ “Detektion gestürzter Personen in häuslicher Einsatzumgebung”. In: *Proceedings Innteract Conference 2016*. 2016, pp. 122_129.
[8] F. Zhao et al. “Real-Time Detection of Fall From Bed Using a Single Depth Camera”. In: *IEEE Transactions on Automation Science and Engineering* (2018), pp. 1_15

Interactive tactile map for visually impaired older adults

Dominika Palivcova

Faculty of Electrical Engineering,
Czech Technical University in Prague, Czech Republic

Abstract. We present our design of interactive tactile map tailored for visually impaired older adults, which employs salient tactile symbols and interactive tactile symbols. Furthermore, we introduced the route-guidance functionality of the map that allows the users to follow and learn particular routes. Qualitative evaluation with representatives of our target audience (mean age was 81.4 years) indicates the ability of our concept to help our audience building spatial knowledge of the indoor environment.

1 Introduction

Aids enabling visually impaired people (VI) to improve their spatial orientation and navigation are of high importance, as visual impairment mainly limits a person's mobility and reduces travel-related activities [2]. According to WHO, the majority of visually impaired is older [3], and many visual limitations are known to be age-related, e.g., myopia, macular degeneration. Older adults differ from the younger ones in their abilities, and in general, the performance of human sensors declines with age. Unfortunately, the current research for visually impaired omits older adults, despite there are significant characteristics of this user group.

For research purposes, we cooperate with a daycare facility for visually impaired (older) adults: Home Palata. It is a complex building comprised of four floors. The clients of Palata are mainly women older than 60 years whose visual impairment varies in severity. Semi structured interviews with 21 clients of Home Palata revealed their problems related to spatial orientation in the building. Some clients have incorrect or missing spatial knowledge of their surrounding environment. Due to this, some clients are afraid to leave their room on their own, which contributes to their social isolation and lowers their independence.

We decided to facilitate the process of familiarization with the new environment by employing the interactive tactile map. The map will be used during the mobility training lessons, and it will also be available to individual clients daily.

2 Design

In the design, we followed the User-Centered Design [1] methodology and created two generations of a tactile map.

2.1 Initial design

In the initial phases, we experimented with different levels of detail of the map and designs of tactile symbols for individual elements that are present in the interior. We used rapid prototyping and 3D printing and evaluated individual designs with a blind expert user. Selected symbols were implemented into the low-fidelity prototype of the map (see Figure 1). In the low-fidelity prototype, the only interactive parts were the room buttons (see Figure 1, d.), which played audio with a description of the room after the user pushed the button.

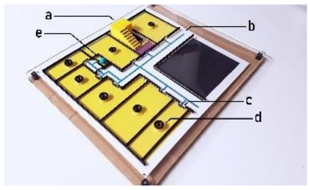


Fig. 1. The initial design of the map. a. Stairs, b. Hallway, c. Doors, d. Interactive symbol for the room, *room buttons* (when pushed, an audio label was played, e.g., 'Culture Room'), e. Symbol for the toilet.

Evaluation. We evaluated this prototype with nine representatives (8 women, average age = 84.4, $SD = 6.6$, $MIN = 70$, $MAX = 90$) of our target audience. Severity of visual impairments of the participants ranged from category 3 to 5 of WHO classification [4] (1 × *cat.*3, 5 × *cat.*4, 3 × *cat.* cat.5), and onset of the impairment varied too. The evaluation was focused on recognizability of individual tactile symbols and qualitative subjective comments on the device

Results. Results of the study showed that chosen designs of tactile symbols are suitable VI older adults, and all the participants praised the audio labels. On the other hand, we received mainly neutral or slightly negative comments on the device in general. Participants *P1, P3, P9* stated they would not explore it because it is too complex for them. There emerged requests for exploration of routes using the map, but at the same time, the map did not offer sufficient support for this use-case, as *P2, P3, P6, P7* were not able to use the map for route-exploration task.

2.2 Advanced prototype

We addressed found issues in the design of the new advanced prototype, which represented artificially created indoor environment containing 11 rooms and three crossings (see Figure 2). We designed a new interaction technique for the map: *route-guidance mode*. It is a non-visual technique that uses touch-detection and audio feedback to guide the user through a route on the map. We employed a special tactile symbol for denotation of the endpoints of the route. Then, we made the area of the hallway in the map touch-sensitive and divided it (virtually) into small segments. Audio feedback plays the most crucial role. As users explore the map from the start point, the map plays increasing tones of piano, if they proceed on the route. When they start to explore the direction that is not included in the route, the map plays a subtle warning tone we call 'no-go' sound.

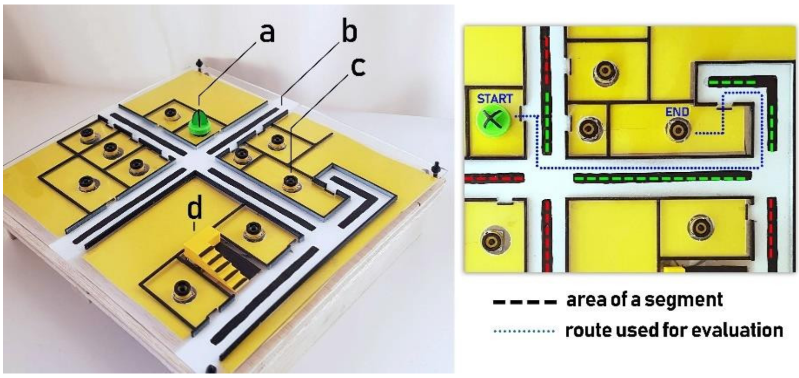


Fig. 2. Advanced prototype with route guidance mode. On the left: a. Symbol for the starting point for navigation, b. Hallway marked by two raised lines and high-contrast, interactive touch segments in the middle of it, c. A symbol for a room and button for triggering an audio description, d. Stairs. On the right: dashed lines denote touch-sensitive segments – green segments played a major scale of piano tones, the red ones played the error tone, dotted line denotes the route used for evaluation, arrow shows the position of the participant towards the map

Evaluation. In the following user study, our goal was to evaluate the *route-guidance mode*. We recruited 10 representatives of our target user group: 7 women, 3 men, average age=81.4 (SD = 11.47, MIN = 52, MAX = 91). The severity of visual impairments varied as well as the onset of the impairment.

We placed the map on the table, and firstly, we let the participants recognize and learn individual symbols. Then we turned *route-guidance mode* on, explained its purpose, and let the participants use the mode to find a particular route (denoted by the dotted line in Figure 2). Once they followed the route correctly, we asked them to learn it. After they stated they are ready, we let them explain the route from the first-person view.

Results. We verified that the concept of *route-guidance mode* is applicable, as it helped seven of nine participants to perform the route-exploration task with the map. Listed participants correctly recognized the sounds provided by the map (i.e., they stopped exploring the direction when 'no-go' sound was played), and they were able to complete the route successfully. The completion times varied from 32 seconds to 242 seconds ($mean$ = 138, SD = 85). Furthermore, the map helped the participants to learn the route. All the participants who completed the route were able to recall it from the egocentric frame of reference. Compared to the results of the first experiment, the participants seemed to be more motivated to use such a device, e.g., *P10* said: *"I think it is usable even in the darkness in which I live now. It would take some time, but I like it, and I would definitely try to use it."*

On the other hand, we revealed new issues in the design. The most common issue experienced by *P6, P7, P10* was caused by employing multiple fingers of the same hand at once for tactile exploration. It sometimes triggered audio feedback multiple times. *Route-guidance mode* was still usable, but the participants perceived that some sounds are redundant.

3 Conclusion

Employing the User-Centered Design approach, we designed an interactive tactile map that uses suitable interaction methods for visually impaired older adults. We presented novel *route-guidance functionality* of the map that seems to help to increase the perceived usefulness of the device and support the acquisition of spatial knowledge.

In future work, we will ensure that the map provides support for touching multiple sensors at once. Also, it would be helpful to conduct a comparative evaluation of our concept to other methods used for the creation of a cognitive model of the environment, e.g., direct exploration, route descriptions, or even another type of tactile map.

4 Acknowledgements

This research has been supported by the project RCI (reg. no. CZ.02.1.01/0.0/0.0/16_019/0000765) supported by the EU and by the project funded by grant no. SGS19/178/OHK3/3T/13 (FIS 13139/161/1611937C0000).

References

[1] DIS, I.: 9241-210: 2010. Ergonomics of human system interaction-part 210: Human-centred design for interactive systems. International Standardization Organization (ISO). Switzerland (2009)

[2] Golledge, R.G.: Geography and the disabled: a survey with special reference to vision impaired and blind populations. Transactions of the Institute of British Geographers pp. 63–85 (1993)

[3] WHO: Global data on visual impairments 2010. Geneva: World Health Organization (2012)

[4] WHO: International classification of diseases, 11th revision (Dec 2018), https://icd.who.int/browse11/l-m/en#/http://id.who.int/icd/entity/1103667651

Physical 3D LED display

Vojtěch Leischner

Faculty of Electrical Engineering,
Czech Technical University in Prague, Czech Republic

Abstract. We have built an active 3D display made from 10500 RGB LEDs arranged in an organic oval shape. The display serves as a canvas for users to draw in 3D space equipped with a special controller. Besides, the 3D drawing capability display also offers stand-alone dynamic light scenes.

1 Introduction

Added interactivity in light installation engages the audience. To make the system intuitive, we rely on a well-known metaphor for painting with a brush. By using the known scenario, we ensure enjoyable user experience. Extensive user testing was deployed to find the right responsiveness of the system. We also make the installation less intrusive and easily maintainable by using a single RGB camera with a visible light spectrum filter as a sensor. In the brush controller, we use infrared light LED as a tracking beacon. Therefore we can reliably filter out any noise that comes from RGB LEDs of the display as well as other ambient artificial lightning.

2 Related Work

This work builds on the existing tracking system. It is based on Infra-red tracking and inspired by Wii Remote Tracking System [1]. It is different in the intended installation - instead of mounting the camera in front of the user, our solution uses ceiling mounting. The proposed solution is also highly modular and scalable and communicates on freely available Open sound protocol to enable uses with other applications. We use only one camera for simplicity and to make sensors as little obtrusive as possible. Depth is estimated from the scale of the remote beacon sphere made out of sandblasted glass.

3 Software structure

We have designed custom PCB modules with programmable LEDs mounted on both sides. We are using an Art-net protocol to control them individually from a central PC. Another PC is analyzing the movement of the tracking beacon and its size to estimate position in 3-D space. These coordinates are sent with OSC protocol back to the central PC.

Fig. 1. Prototype at Euroluce Milan 2019

4 Conclusion and Future Work

We have successfully built the active LED 3D display and tested it in the 2019 Euroluce Milan Light Fair. 10500LEDs proved to be a challenge for an individual controlling system in the means of network bandwidth and computer processing power as well as suppressing noise generated by current induced by coils created by cabling. Although successful the installation prompted the research into other types of 3D displays, particularly in passive surfaces lighted by a single programmable light source.

Another useful feedback for future work is to design dynamic user profiles for interactive interfaces. Different people have different body features such as length of arms and height that affects the responsiveness of the system. Therefore specific user profiles should be automatically created for each user to account for individual properties.

Acknowledgements The project has been funded by grant no. SGS19/178/OH K3/3T/13 (FIS 13139/161/1611937C000).

References

[1] van den Heuvel, J., Bikker, J.: Detecting 3d position and orientation of a wii remote using webcams. In: Allbeck, J.M., Faloutsos, P. (eds.) Motion in Games. pp. 448–458. Springer Berlin Heidelberg, Berlin, Heidelberg (2011)