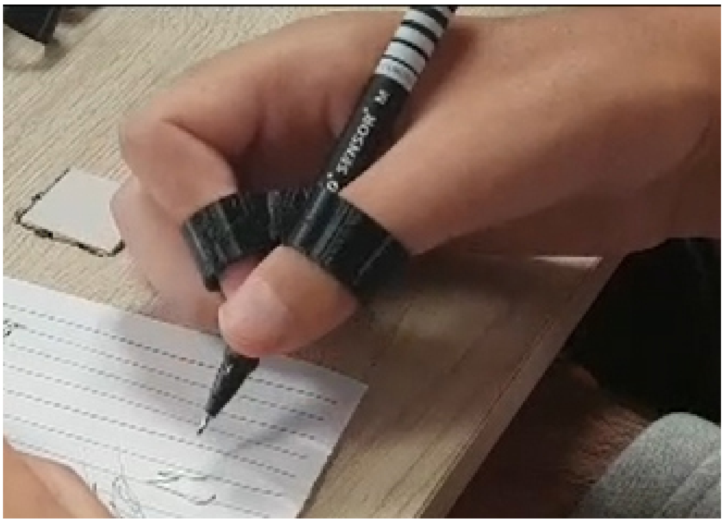
Promotion 2024 - **4e année cycle ingénieur**
Majeure **Creative Technology**

INRIA
Groupe WILLOW
Paris, France

Reconstruction de la forme d'une main pour génération automatique d'un modèle 3D

d'une orthèse d'écriture

Hippolyte Bonabeau



Thanks

I'd like to express my profound gratitude to my tutor, Etienne Moullet, whose guidance and expertise were pivotal in the realization of this project. Working alongside a diverse and talented team made up of research engineers, esteemed researchers, dedicated post-doctorates, insightful thesis students, and fellow interns like Gabriel, Eugene, and Matthieu has been an invaluable experience. The synergy, knowledge, and support from each member played a crucial role in my learning and the project's success. My heartfelt thanks to everyone for their unwavering encouragement and for making this internship a truly transformative learning journey.

Table of Contents

1 - Introduction

During my internship at Inria [1], a research institute in computer science and mathematics, I was introduced to the Humanlab-Inria [9] initiative. This project is driven by a commitment to addressing the unique needs of individuals with disabilities. Specifically, in collaboration with the Humanlab Saint-Pierre de Palavas, there was a focus on devising solutions that leverage technology to aid those with hand mobility issues.

The main challenge was that many of the existing writing orthoses, devices designed to aid individuals with limited hand mobility, were crafted through cumbersome, imprecise, and hard-to-replicate processes. While there was a specific CAD model for an orthosis available, it was tailored for my tutor, Etienne Moullet with his own measurements and lacked adaptability for a broader audience. The design and measurements were found through trial and error, thanks to 3D printing's rapid prototyping nature. This realization highlighted a gap in providing a more generalized, yet tailored, solution for a diverse set of users. This solution could be generalized by using the original design and tweaking key measurements and angles.

My core responsibility revolved around designing a software solution capable of scanning a human hand to determine its measurements and subsequently generate a tailored 3D orthosis model. This involved harnessing recent advancements in AI-based computer vision tools to convert hand images or video streams into a 3D mesh. The objective was not just to create an ergonomic design, but also to simplify the measurement phase, thus making the entire process more user-friendly and replicable.

After a period of 4 months of research and collaboration with diverse teams, including computer vision experts, the team from from Humanlab Saint-Pierre de Palavas, and my tutor, the project holder, I successfully developed a software tool that met our objectives. The tool not only streamlined the process of creating orthoses but also made it possible for other users to adapt the designs for their unique needs. The developed application regroups the calibration, picture-taking and orthoses rendering processes. The culmination of this project was the documentation of the solution on a github repository [2], making it accessible and setting the stage for future innovations in the field.

2 – Position at INRIA

2.1 - Actors

2.1.a - INRIA (*Institut national de recherche en informatique et en automatique*)

Inria stands out as a premier research institution dedicated to the intricacies of computer sciences, digital technologies, and mathematics. Unlike many organizations that diversify their attention across a myriad of domains, Inria's focus remains steadfast, spanning from artificial intelligence to cybersecurity and robotics. The depth of their commitment ensures groundbreaking discoveries and solutions in their chosen fields of expertise.

An exemplary manifestation of Inria's real-world applications is their Humanlab-Inria initiative. This initiative highlights their unwavering commitment to translating high-end research into tangible impacts. Specifically, the initiative aims to harness technological advancements to aid individuals with disabilities, emphasizing Inria’s ability to interlink research, innovation, and societal welfare.

Collaboration remains at the heart of Inria's operational strategy. Since its inception in 1967, Inria has actively sought partnerships with varied entities—be it esteemed academic institutions, industry behemoths, or nascent start-ups. Such partnerships not only ensure a multifaceted approach to research and development but also offer a synergy that brings forth diverse perspectives, enriching the research process and its subsequent applications.

Inria's strategic foresight extends to the ethical implications of technological and mathematical solutions. With over 200 research teams across France and a commitment to train young researchers, their methodologies are intricately designed to incorporate ethical considerations, ensuring that the advancements they champion are not just innovative but also conscientious.

In conclusion, Inria's unique positioning within its sector is a confluence of deep-focused research, real-world applications, open-source advocacy, collaborative endeavors, and ethical considerations. Their holistic approach, coupled with their history of engagement with global stakeholders, sets them apart, affirming their role as a trendsetter in the world of computer sciences, digital technologies, and mathematics.

2.1.b - HUMANLAB INRIA

"My Human Kit" [11] orchestrates and develops the "Humanlab Network", a collective of venues where individuals collaborate to design, create, and share technical aids for disabilities. Since the inception of the first humanlab in the Rennes region in 2017, others have sprung up throughout France and internationally.

The Inria's humanlab consists of researchers and research engineers based in Grenoble and Montpellier. It is dedicated to addressing the needs voiced by people with disabilities within the Humanlab network, facilitated by allocating work hours from its researchers and engineers. Weekly meetings between teams took place to enable and promote discusson and collaboration.

2.2 - Role and Position

At INRIA, I assumed the role of an intern with the primary objective of bridging the gap between technological advancements and their practical applications for individuals with disabilities. Under the mentorship of my tutor, Etienne Moullet, I was entrusted with the task of transforming an idea into a tangible software solution. While the foundation was an existing CAD model for an orthosis tailored specifically to my tutor, the challenge lay in its adaptability and scalability. My position necessitated a deep dive into AI-based computer vision and an understanding of 3D modeling intricacies. Collaborating weekly with the computer vision experts and the dedicated team from Humanlab Saint-Pierre de Palavas, I played an instrumental role in translating research and insights into a user-friendly tool. This immersion into the project allowed me not just to apply my academic knowledge but also to shape a vision into a replicable solution, solidifying my understanding of the iterative and interdisciplinary nature of tech-based solutions.

3 . Presentation of internship

3.1 – Projects and objectives

3.1.a - Inkredable Hand Mesh Reconstruction

Introduction

Harnessing Microsoft's open-source mesh reconstruction repository, Mesh Graphormer [3], the Inkredable project generates 3D hand meshes from centered images of hands. With these meshes, the tool reconstructs the spatial positioning of the user's hand, which is then employed to derive the unique dimensions of the fingers. This innovation culminates in the generation of a 3D-printable STL orthosis file specifically designed for writing.

This project's intent was twofold. First, it aimed to be an open-source solution that could be easily accessed and used by anyone with a simple RGB webcam and a reference point, promoting inclusivity. Second, it sought to serve individuals who face challenges in writing, including those with disabilities, by allowing them to 3D print a tailor-made orthosis, further emphasizing the project's commitment to creating assistive solutions.

Planning

May :

i. Define research objectives and gather relevant papers and theses.
ii. Gather relevant models/tools to be used and repurposed.
iii. Dive into research papers and test open-source tools.
iv. Discuss all findings and choose most relevant and useful solution, that can be scaled and modulated to our needs.

June:

i. Test different techniques for depth perception (monocular depth estimation, reference object).
ii. Create simple UI for testing, test different techniques for finding the correct scale factor (reprojection error, 2d landmark detection on mesh renders).
iii. Build simple pipeline with hand/coin detection and inference and test hand positions and situations.

July:

i. Refine pipeline, start developing backend with flask.
ii. Develop and refine front end user experience.
iii. Organize and comment code, adding automatic orthosis generation.

August:

i. Add docker to the project for faster and easier deployment.
ii. Add documentation and installation instructions.

Needs

Must Have:

- Hand and Coin Detection: Essential for core functionality and accurate measurement and depth retrieval.
- Method for Depth Perception: To recover the depth value of the hand at varying distances, crucial for the project's precision.
- Generation of 3D Hand Mesh: Transforming 2D images into accurate 3D representations for precise orthosis creation with measurements.
- Retrieval of circumference of key fingers of hand: Ensuring measurements are accurate and consistent for both hands, fundamental for the custom orthosis.

Should Have:

- UI for Capturing Pictures: An intuitive interface is essential for users to easily provide input images.
- UI for Calibrating Camera: For users to ensure accurate depth and dimension perception through their own devices.
- Way of Generating STL File with Given Measurements: To transform retrieved data into a tangible 3D-printable format.

Could Have:

- Web application for integration of end to end pipeline and ease of use.
- *Regenerating Orthosis Shape based on most comfortable writing position for user: A user-centric feature that can provide a more personalized and comfortable fit.*
- *Deployed Website for Immediate Use: To provide users with a platform that is easily accessible and ready for immediate interaction.*
- *App deployment on mobile: enable users to easily take pictures and calibrate, anywhere and anytime.*

Note: The bullet points in italics have not yet been integrated

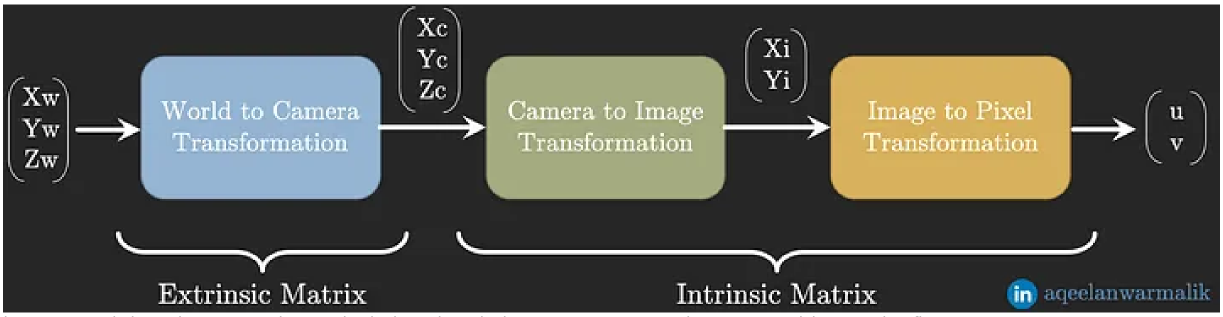
Methods and tools

Projet = mesurer
contrainte = monocualar
outils dispo: nn models, - computer vision

Wether its pose estimation or object recognition, with using a single camera frame one of the biggest challenges is discerning if an object is small and close to the camera or far away and very large. Models that perform pose estimation on hands approximate depth using average hand sizes learned during training.

Depth estimation can be very difficult without a reference object with known size, and With only one camera to be used, the depth of the hand could be estimated using monocular depth estimation, .

Camera calibration using OpenCV becomes crucial in accurately estimating depth. In my project, I employed OpenCV to perform camera calibration, which involved determining the intrinsic parameters of the camera such as focal length, principal point, and distortion coefficients. These intrinsic parameters play a pivotal role in creating a camera matrix, which serves as the foundation for accurately projecting 3D world coordinates onto the 2D image plane.



https://towardsdatascience.com/what-are-intrinsic-and-extrinsic-camera-parameters-in-computer-vision-7071b72fb8ec

With the camera calibrated, I extended my application to detect a 2 euro coin and a hand in the scene. Leveraging the intrinsic camera parameters, I computed the depth of the 2 euro coin using the formula: Depth =Known Coin Diameter × Focal Length /Apparent Coin Diameter.

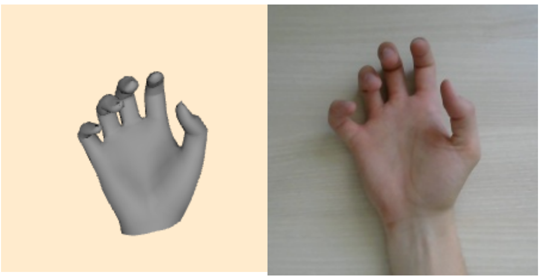
This formula allowed me to determine the distance of the coin from the camera, providing a more precise depth estimation than typical monocular methods. Using the vertices generated by the Mesh Graphormer model, the intrinsic camera parameters and the given depth, I rendered a normalized mesh and placed it in corresponding 3D space.

Alongside Mesh Graphormer, the MANO [4] hand model, a parametric hand model that accurately represents the 3D pose and shape of a human hand is utilized. MANO provides a detailed representation of hand articulation by modeling each joint's degrees of freedom, along with pose parameters that describe the flexion and abduction angles of the fingers

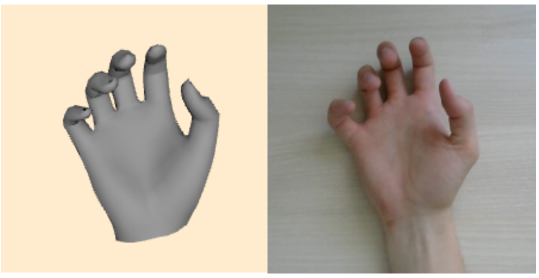
After placing the hand at the correct depth, another challenge arises: finding the correct scale factor by which to multiply the vertices of the mesh to recreate the accurate scale and depth of the hand in the scene, and translating the rendered hand on the x and y axes. To address this, Google's MediaPipe [5] is employed to detect 2D landmarks on both the cropped image of the actual hand and the image of the rendered hand. By comparing the positions of these landmarks, an error metric is computed. Specifically, the difference between the MediaPipe landmark coordinates of the rendered mesh and the landmark coordinates of the real-world hand image serves as the error. This difference, often represented as the sum of squared distances between corresponding landmarks, is the function that the optimization process seeks to minimize.

The mesh is first scaled, then translated. An initial guess for the scaling factor is derived based on the hand's posture, discerned from the angles at key joints using 3D world coordinates provided by MediaPipe. If the angles suggest a closed hand, the initial scaling guess might be reduced since a closed hand typically occupies a more compact space compared to an open one. To achieve the necessary translation, the u,v coordinates from the real-world image of the hand are converted into x,y coordinates using the intrinsinc camera matrix and the depth. This conversion helps in determining the exact amount and direction by which the rendered mesh needs to be shifted. The mesh is then translated accordingly in the scene.

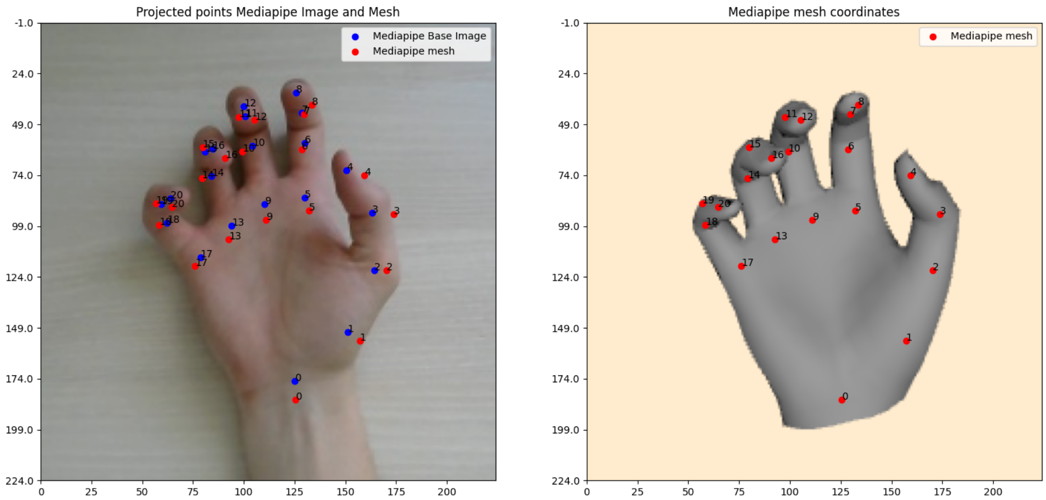
Through this iterative process, which combines both scaling and translational adjustments, the rendered mesh can be aligned more accurately with the real-world image, ensuring the correct positioning and scaling of the hand in the virtual scene.



First render with correct depth and corresponding camera matrix, but incorrect x/y translation and scale

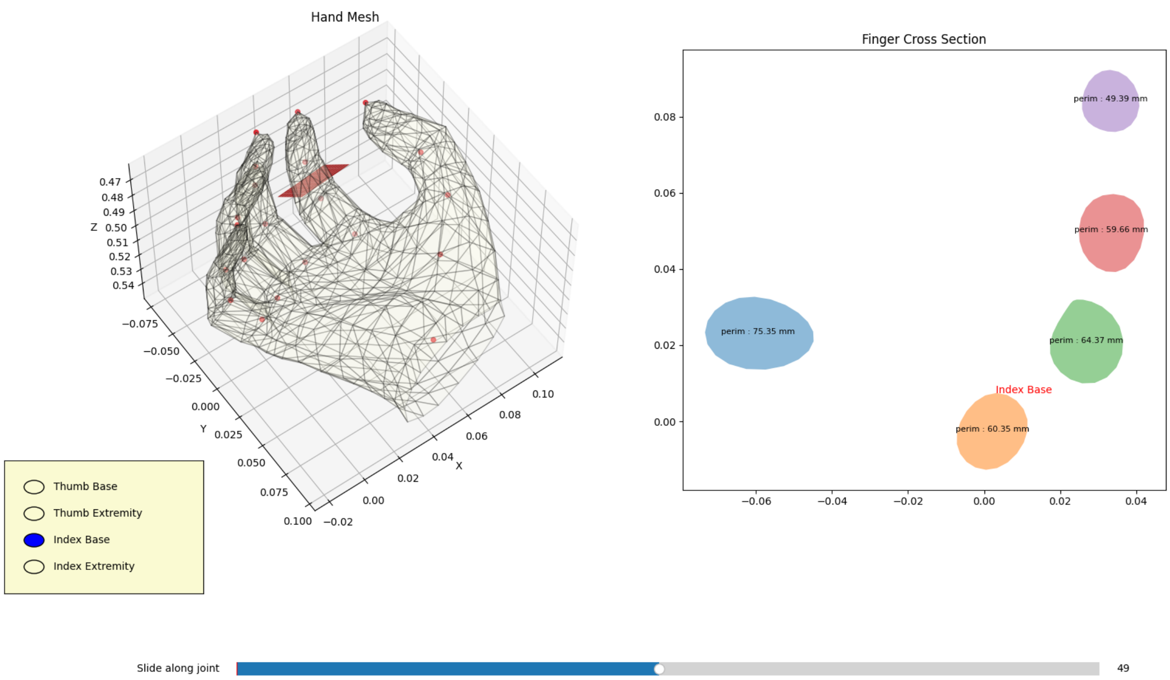


Final render after error minimizing function



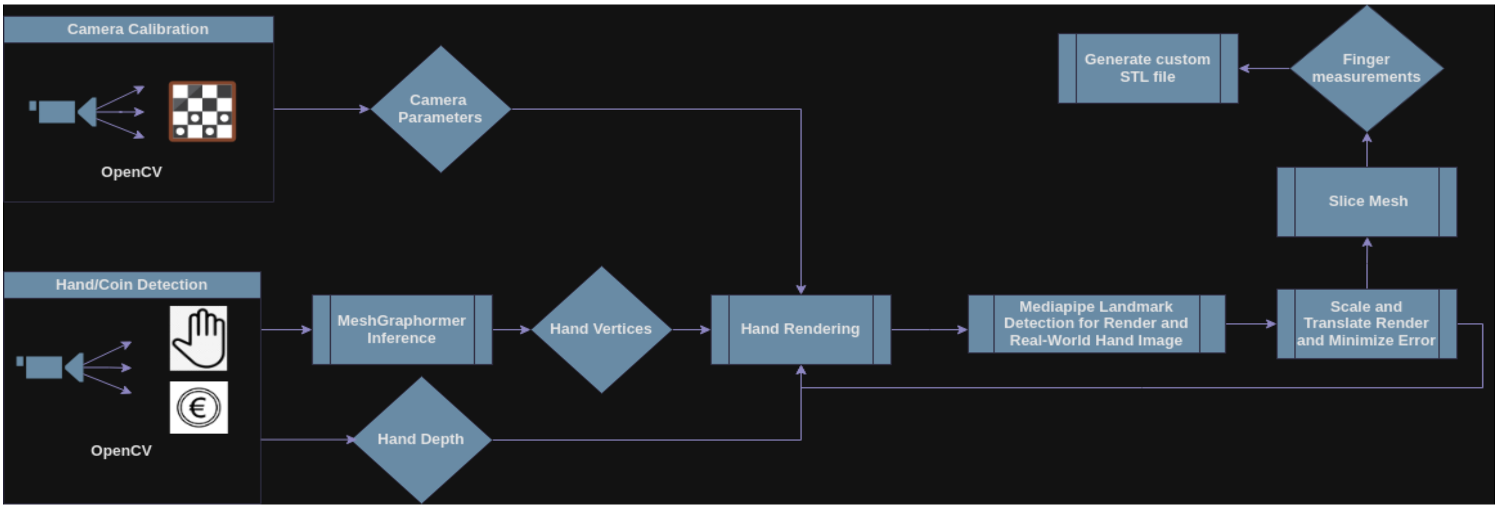
Final result with the red points being the mediapipe landmarks from the mesh render, and the blue points being the real-world hand image.

Once the error function is minimized, we slice the hand mesh and retrieve key hand measurements, including measurements from the index and thumb.



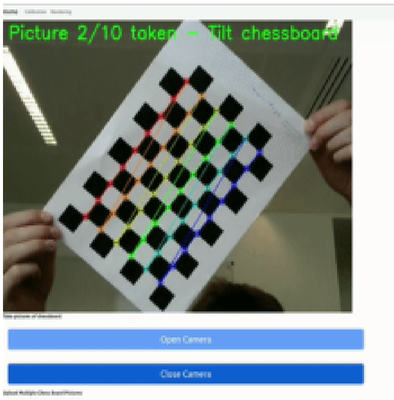
Hand Mesh being sliced along the red plane (left), cross section with the perimeter of each finger (right)

Project Architecture

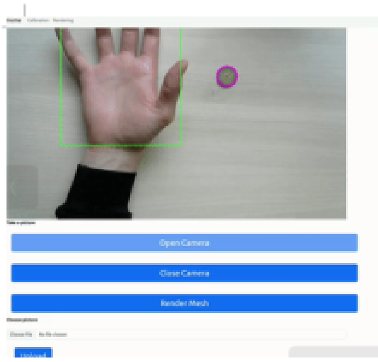


In terms of user accessibility, the project also incorporates a Flask [6] web application. The web app guides users through the camera calibration process, hand and coin detection, 3D mesh rendering, and ultimately the generation of a custom STL file for the orthosis, which can be downloaded and 3D printed. The solution to generate

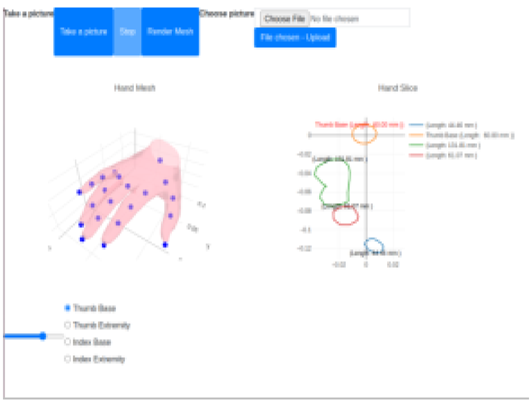
The STL file measurements can be manually tweaked for a better fit.



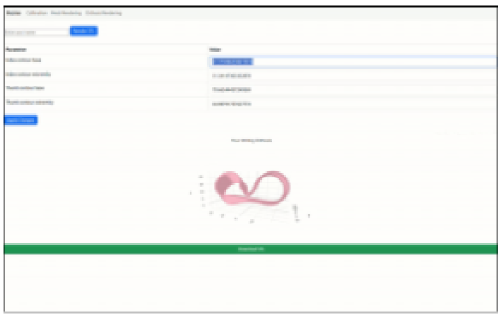
Calibration process



Hand/Coin Detection



Hand Slicing



Orthosis Rendering

Conclusion

The Inkredable project embodies the spirit of democratized technology. By leveraging the commonplace tools of a computer and webcam, it ensures that its innovative solutions are accessible to a vast audience. The open-source foundation further cements its commitment to inclusivity, allowing for a broad swath of users and developers to interact, enhance, and benefit from its offerings.

On a more intimate note, this journey has been profoundly educative. Collaborating one-on-one with my tutor, Etienne Moullet, provided a deeply enriching experience. Together, we navigated the intricacies of the project, and each challenge surmounted augmented my understanding and appreciation for the field. The Inkredable project is a testament to the mix of passion, expertise, and collaboration I had the incredible chance of being a part of.

3.1.b - TOM France – 2-day Hackathon

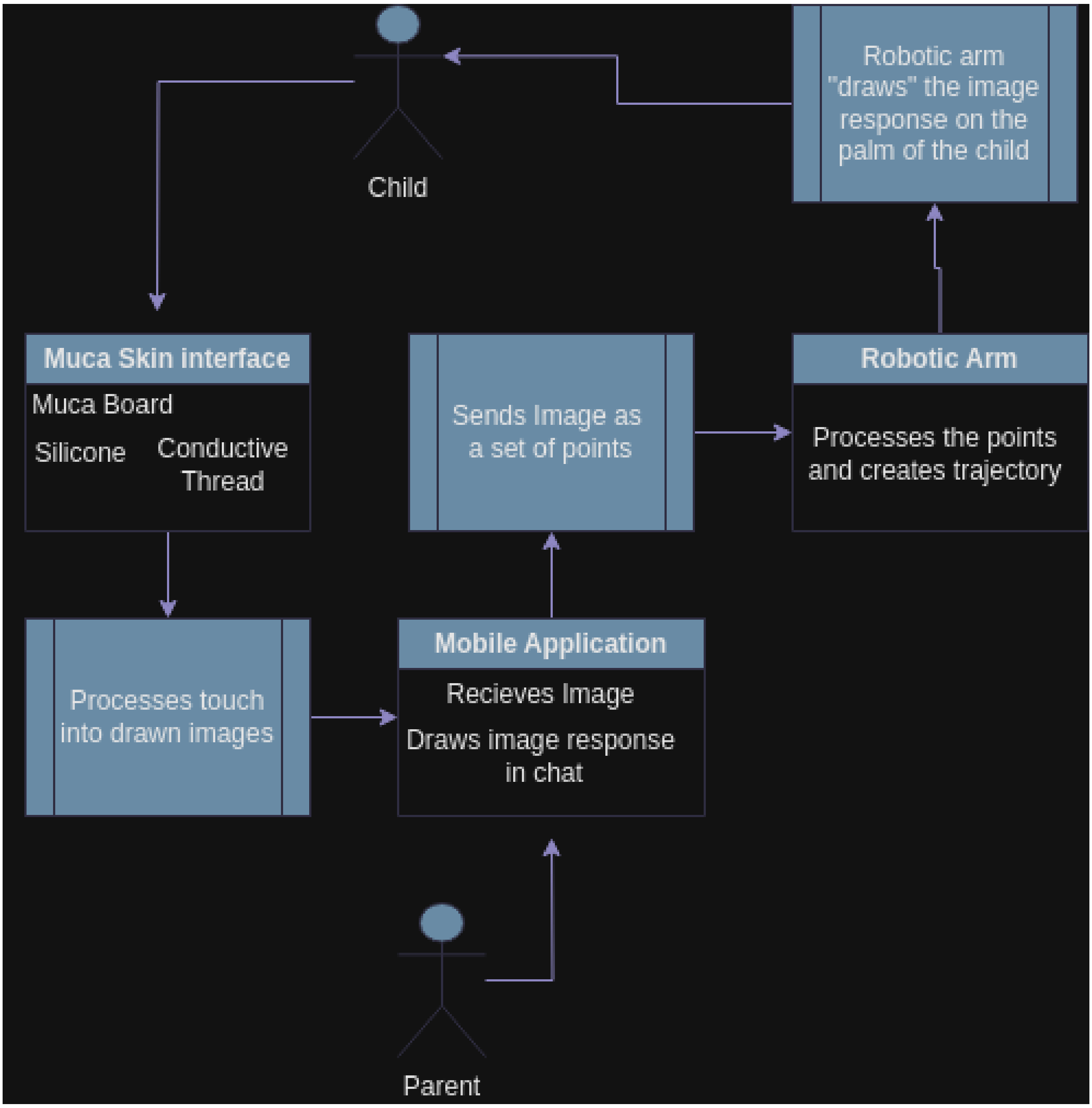
Jean Bouissou, responsible for deafblindness at the Ministry of Solidarity and Health and a deafblind individual himself, recognized the profound challenges faced by those with combined hearing and vision impairments. Deafblindness, a unique and complex condition, intensifies the sense of isolation and creates barriers to communication and access to information. Both sound and sight are either limited or absent for deafblind people, compelling individuals to rely heavily on touch for communication and comprehension.

A solution was envisioned where deafblind children could bridge the communication gap with their parents, even from a distance. The core idea was to establish a system where a child could convey a haptic message, utilizing the tactile sense, to their parents. This would then allow the parents to perceive the sentiment, emotion, or need being communicated and respond similarly, creating a two-way haptic communication channel. Such an initiative would not only provide these children with a sense of connection and belonging but would also empower them with the confidence to communicate and express themselves more freely.

The 2-day hackathon [7] brought together 13 teams composed of engineers, electricians, designers and most importantly a project leader who’s problem was to be solved. Our team had different hubs: UI/UX, design, tactile touch and mechanical engineering. Combining brainstorming, fast prototyping and an extreme hands-on approach, we managed to develop a prototype in the end.

The full project was composed of 3 parts:

- Skin matrix interface using Muca development board [8]
- IOS application that receives haptic messages in the form of images from the skin interface, and can send messages to a robotic arm.
- Robotic pantograph that reproduces the form of the image received from the parents’ applicatio



Project Architecture

Needs

Must Have:

- Tactile/Haptic Interface

This is the primary point of contact for the deafblind child, enabling them to input tactile messages. It's the foundation of the communication system, playing a vital role in the reception and transmission of haptic signals.

- Way of interaction for parents:

A component that receives and decodes the tactile messages from the skin interface. It's essential for ensuring the smooth transition of messages between the child and parents.

- Robotic Arm:

A tool that translates the haptic messages into tangible forms. It reproduces the image received from the parent's application, ensuring that the child can perceive the feedback in a tactile manner, thereby closing the communication loop.

Should Have:

- User-friendly UI for the Application:

The usability of the application is paramount. An intuitive interface would enable parents to seamlessly receive and respond to messages, fostering smooth communication.

- Robust Calibration Mechanism for the Skin Matrix:

To ensure the haptic messages are accurate and consistent.

- Safety Mechanisms for Robotic Pantograph:

Given its interaction with the child, the robotic arm must be safe and gentle.

Could Have:

- Customizable Tactile Patterns for Individual Users:

Allowing both parents and children to create and save specific tactile patterns that hold personal significance can further enrich the communication experience.

- Integration with Other Platforms:

While the focus is on an IOS application, expanding the software's compatibility can make the solution more inclusive and accessible.

- Feedback Mechanism within the App:

An inbuilt system where parents can provide feedback or suggestions based on their usage, aiding in further refinement and improvement of the system.

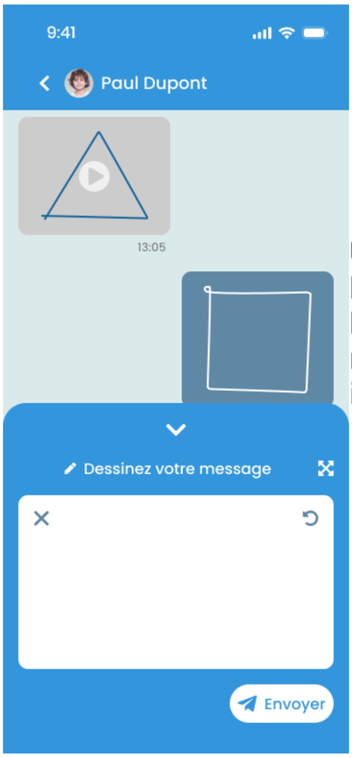
- Training Modules:

Given the unique nature of the communication, having training sessions or tutorials within the app can guide users, especially parents, on how best to utilize the system for effective communication.

Methods and Tools

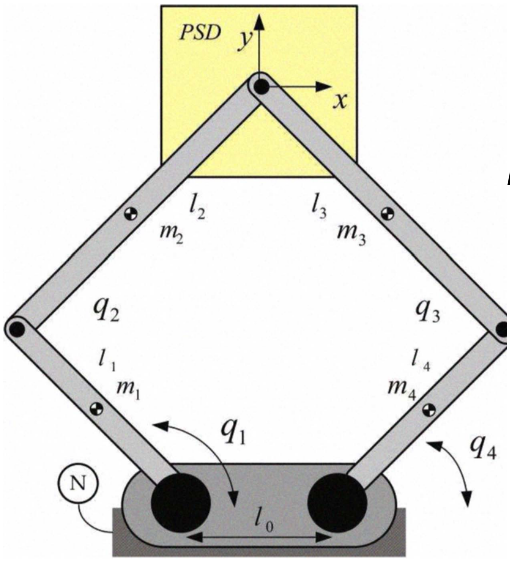
Muca skin interface
https://muca.cc/

The 1st step of the pipeline of the project is a haptic skin interface made of pouring silicone over conductive threads. The multi-touch interface uses project mutual capacitance which identifies touch by measuring the local change in capacitance on the grid array of conductive traces. The artificial skin/multi-touch surface can then be placed as a band on the wrist of the child, connected to a development board and an ESP32 that sends the data over wifi to the parent application. The code is written in Arduino for processing and Processing for visualization.



The fully-functional IOS application was coded by the UI/UX design and development team during the hackathon. The goal was to create a user-friendly chat-based messaging system that had user profiles, and video recordings of the images being drawn on the multi-touch interface. It could then send drawings to the robotic arm.

Parent Application



Pantograph robot example

https://www.researchgate.net/figure/Pantograph-configuration_fig2_263550587

Finally, the pantograph robot was designed from scratch, using an Arduino uno development board that controlled two servo motors at the base. A prototype was made using popsicle sticks, and the final product consisted of a 3d printed base, arms, attached to a wooden base plate. The robot receives a set of points from the drawn image, and moves the servo motors at a certain angle to reach a certain position on the x,y axes. The option of adding a z axis was discussed, but the team did not have enough time to fully prototype the feature. The z axis could be added thanks to the inferred velocity either from the haptic multi-touch skin interface, or could eventually convey the information of the finger being lifted up or down in the application.

Conclusion

The TOM France Hackathon aimed to tackle a profound challenge: breaking the barriers of communication for deafblind individuals. This condition, as described by Jean Bouissou, intensifies feelings of isolation due to the simultaneous loss of both vision and hearing. The primary ambition was to empower deafblind children to communicate with their parents, even at a distance, using a haptic communication system.

Throughout a rigorous two-day session, 13 teams, bursting with enthusiasm and diverse expertise, gathered to innovate. Among them, our team's synergy between UI/UX, design, tactile touch, and mechanical engineering was palpable. The intense period of brainstorming, rapid prototyping, and hands-on development led to the birth of a promising prototype. This solution was a three part system: a skin matrix interface developed on the Muca board, an IOS application that interpreted haptic messages into images, and a robotic pantograph that turned the images back into tactile experiences.

At the core of the system was the tactile interface, acting as the child's voice, enabling them to convey their sentiments. The IOS application served as the bridge, capturing the tactile imprints and translating them for the parents. The robotic pantograph, thoughtfully designed with a combination of 3D printing and popsicle sticks, recreated the messages for tactile interpretation. Safety and usability were prime considerations, ensuring smooth and accident-free communication.

The uniqueness of this project lay not just in its innovative solution, but in the empathy it was built upon. Each component was crafted keeping in mind the users' special needs, from the skin's sensitivity to the application's user-friendliness. The potential to customize tactile patterns and integrate with other platforms, as well as the idea of providing training modules, reveals the team's holistic vision for the project.

While navigating the constraints of a two-day hackathon, particularly the challenge of integrating three distinct projects in such a short time frame, I gained invaluable insights into the collaborative spirit of diverse professions. This experience illuminated not only how I can learn immensely from others but also how I can offer significant contributions to a collective endeavor. Embarking on my first hackathon, I stepped in with an open mind, with no particular expectations. Reflecting on it now, I wholeheartedly encourage others to immerse themselves in the experience that TOM France offers.

4 - Feedback

During the four months of my internship, I gained a variety of skills, both technical and organizational. In addition to my main tasks, I had the opportunity to participate in other initiatives, such as a hackathon directly connected to my internship. This event focused on finding innovative solutions for people with disabilities, giving me a broader perspective on the impact of technology in real-world scenarios.

Technical skills :

- Python Programming:

The backbone of the Inkredable project relies heavily on Python, with sporadic usage of HTML and JavaScript to underpin the web application component. Engaging with this diverse technology stack facilitated a deeper comprehension of Python and provided an avenue to harness powerful libraries like pyrender, trimesh, and scipy.

- Neural Networks:

Central to the Inkredable project is the MeshGraphormer model. It uniquely combines elements from graph-based and transformer neural networks, supplemented with specialized modules. By immersing myself in the intricacies of the codebase, I cultivated a robust understanding of these neural architectures, gaining the expertise to adapt and refine them further.

- Computer Vision:

Being enveloped in a project so deeply rooted in computer vision has been an enriching experience. It necessitated a thorough exploration of quintessential computer vision concepts such as camera calibrations, sophisticated rendering techniques, pose estimation, and more. This venture equipped me with hands-on skills and knowledge crucial to the domain of computer vision.

Other skills :

Thanks to reading many papers and a thesis during the early stages of my internship, I learned how to efficiently read and better understand scientific papers. I also learned the intricacies of how to build a well organized python project, as this was the biggest project I have worked on to date. It was important to clean up and organize the code to make it readable, understandable, and so more importantly reusable for the future. During this project, working one-on-one with my tutor provided a unique landscape of both challenges and opportunities. Given the absence of preset deadlines, I took the initiative to craft my own planning and set a broad timeline, anchoring it around the four months I had allocated for the project. This not only taught me the discipline of self-regulation but also enhanced my skills in project management and foresight. Regularly presenting my progress to my tutor, I honed my ability to succinctly and lucidly convey complex ideas. This consistent interaction, combined with the responsibility of setting my own milestones, underscored the significance of autonomy in research.

Creative Technology

During my time in the Creative Technology major [10], I was introduced to a kaleidoscope of disciplines, honing a multifaceted approach that fuses technological advancements with philosophical and critical frameworks. This robust foundation not only heightened my insights into potential user needs, stemming from my exploration into project marketing and crowdfunding, but it also equipped me to navigate the intricate challenges I encountered during my internship at INRIA in the willow team particularly for the Inkredable project. The project's essence, stands as a testament to the symbiosis of human-computer interactions and AI, integral facets of my academic program.

The tangible experience I acquired in mechanics, electronics, and emerging materials became instrumental in morphing abstract concepts into real-world applications. Such a hands-on approach was further amplified by my deep-seated understanding of artificial intelligence, mixed reality, and the nuances of biotechnological integrations. My endeavors in advanced programming and UX design, paralleled with the principles of large-scale industrialization and complex software architecture, refined my proficiency to optimize and scale projects, ensuring they resonate with end-users while achieving operational efficiency.

One of the pinnacle achievements, the conversion of RGB webcam inputs into 3D-printable orthosis files, reflects the major's ethos of radical learning, antidisciplinarity, and collective intelligence. It signifies the bridge I built between my academic learnings and practical applications. Furthermore, my ability to orchestrate harmony between diverse units like R&D, marketing, and production became a linchpin for smooth project transitions. Consequently, with the bedrock knowledge and skills garnered from Creative Technology, my contributions at INRIA have been instrumental in pushing the boundaries of technological innovation, while always anchoring on a profound human impact.

Bibliography

[1] https://www.inria.fr/fr

[2] https://github.com/hippobo/Inkreadable-Hand-Mesh-Reconstruction

[3] https://github.com/microsoft/MeshGraphormer

[4] https://mano.is.tue.mpg.de/

[5] https://developers.google.com/mediapipe

[6] https://flask.palletsprojects.com/en/2.3.x/

[7] https://www.tomfrance.fr/

[8] https://muca.cc/

[9] https://project.inria.fr/humanlabinria/fr/

[10] https://dvic.devinci.fr/

[11] https://myhumankit.org/