Northwestern University

McCormick School of Engineering

Department of Electrical Engineering and Computer Science

CE 347 – Microprocessor Systems Project II

Spring Quarter - 06/01/2024

AI Smart Mirror | [Github]

Design Document

Juin Chen Oh - juinoh2024@u.northwestern.edu

Youssef Ibrahim - youssefibrahim2024@u.northwestern.edu

Table of contents

1. Introduction

The AI Smart Mirror is a cutting-edge product designed to revolutionize the way individuals interact with their wardrobe. Leveraging advanced AI technology, our smart mirror provides users with personalized fashion advice and virtual outfit try-ons, enhancing their daily routine and boosting their confidence. This innovative solution combines high-resolution imaging, sophisticated body measurement algorithms, and state-of-the-art outfit visualization to offer an unparalleled user experience. Our product aims to make fashion choices more accessible, accurate, and enjoyable.

The AI Smart Mirror operates through a seamless integration of hardware and software components. Using a 75-inch TV as a reflective surface, the mirror is equipped with a Jetson Nano running Linux and a 4K camera. The software utilizes the MediaPipe Pose module to accurately measure various body dimensions, achieving a remarkable 95% accuracy rate. Following this, users can browse a curated list of outfits and visualize themselves in different styles using the OOTD model powered by Stable Diffusion. This not only enhances the shopping experience but also promotes better wardrobe management and personalized style recommendations.

2. Overview

Our product consists of two main components: a hardware setup and an advanced software system. The hardware setup includes a 75-inch TV screen that functions as a mirror, a Jetson Orin processing unit, and a 4K camera. This configuration captures high-quality images of the user, which are essential for accurate body measurement and outfit visualization.

The software component comprises several key modules:

1- **Image Capture and Processing:** The 4K camera captures images of the user, which are then processed by the Jetson Nano.

2- **Body Measurement Analysis:** Utilizing the MediaPipe Pose module, the system measures key body dimensions such as shoulder width, chest circumference, waist size, and height. This module ensures high accuracy, providing reliable measurements 95% of the time.

3- **Outfit Selection Interface:** Users can interact with a user-friendly interface to browse and select outfits. This interface is designed to be intuitive, making it easy for users to explore various fashion options.

4- **Virtual Try-On Feature:** Using the OOTD model with Stable Diffusion, the system superimposes selected outfits onto the user's image, offering a realistic preview of how the clothing will look on them. This feature enhances the shopping experience by allowing users to see themselves in different outfits before making a purchase.

3. System Design Analysis

3.1 High Level System Design and Operation

Our final design consists of two major components that communicate over a network connection: a client application and a server application. The client has three key components: a 75-inch TV that acts as a mirror, a Jetson Orin running Linux, and a 4K camera. The server application runs the OOTDiffusion AI model to provide virtual try-on services for the client.

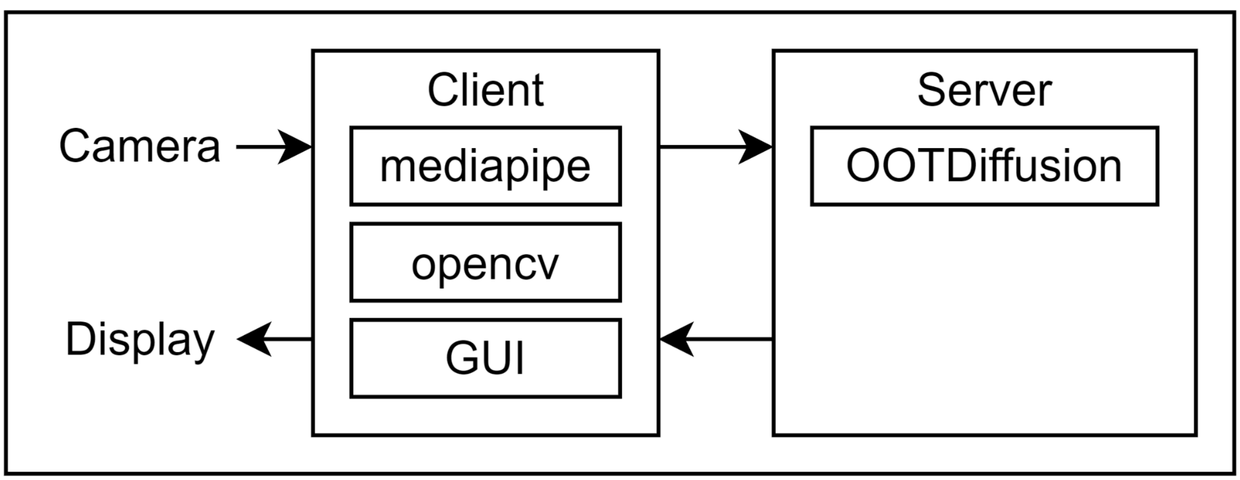

Figure 1: System Block Diagram

The client application is the component that the user interacts with. It is responsible for taking a picture of the user, calculating the body measurements of the user, and allowing the user to select a top and bottom that they would like to try. The client also cleans up the image of the user by using image segmentation models to remove the background of the image. This process enhances the performance of OOTDiffusion.

Once the image of the user is ready, the client uploads the image, along with the user's choice of top and bottom to the server application. The server is only responsible for one thing, which is to run the OOTDiffusion model that puts clothes onto the image of the user. The reason a server-client architecture was necessary is that we were not able to get OOTDiffusion to run on the Jetson Orin. We are still not sure why it doesn't run, but we suspect that it is because OOTDiffusion is dependent on at least one Python library that is not available on ARM. Once OOTDiffusion generates the virtual try-on image, the server sends the image to the client, which then displays it for the user to see.

3.2 Design Rationale

Our team chose a hardware module centered around a TV screen for its multifunctional use as both a mirror and a display. This decision was based on a thorough analysis of user needs and preferences, ensuring that our design is user-centric, affordable, and accessible.

Key Design Considerations:

User Comfort and Accessibility: The large TV screen provides a clear and comprehensive view, making it accessible to users of all ages and abilities. The use of familiar technology (a TV) also helps in minimizing the learning curve for new users.

High-Quality Imaging: The 4K camera ensures that the images captured are of high quality, which is essential for accurate body measurements and realistic virtual try-ons.

Powerful Processing: The Jetson Nano was chosen for its ability to handle complex AI algorithms efficiently, ensuring quick and accurate processing of images and measurements.

Seamless User Experience: The user interface is designed to be intuitive, with simple navigation for selecting outfits and viewing virtual try-ons. This ease of use is crucial for encouraging regular use of the product.

Privacy and Security: All image processing is done locally on the device, ensuring that user data remains private and secure. No data is transmitted externally without user consent.

By focusing on these design principles, we aimed to create a product that not only meets the functional requirements but also provides a comfortable and enjoyable user experience. The decision to use a large TV screen, a high-resolution camera, and powerful processing hardware ensures that the AI Smart Mirror is both effective and user-friendly.

4. Ethical Considerations

Like any other technological advancement, the AI Smart Mirror raises ethical considerations regarding data privacy, data security, and potential technological misuse that must be addressed to ensure it remains a largely beneficial product for its intended users and beyond. We plan to implement these considerations in our final product iteration.

To mitigate the primary concerns regarding data privacy and security, we will implement robust data privacy measures in data transmission and authentication measures for data usage to prevent unauthorized interception of the user's sensitive information. Such measures will include encrypting data transmissions between the AI Smart Mirror and any connected devices, obtaining explicit user consent regarding data usage, applying secure authentication methods such as one-time passcodes before application access, and anonymizing captured images before any data transmission.

We will strictly follow data protection regulations like the General Data Protection Regulation (GDPR) to ensure we collect, process, and store user data responsibly. In our implementation, we will obtain informed user consent before collecting personal data, be transparent about how their data is collected and used, and allow users to control how long their data is stored and when to delete it.

To combat potential misuse of the technology through unauthorized access, we will regularly update the application to patch security vulnerabilities and conduct thorough security audits to identify and address potential system weaknesses. We will also introduce access controls to limit who can manipulate and access sensitive data.

Addressing the ethical concerns surrounding recording people and utilizing ML models for human classification tasks, our team will emphasize the importance of obtaining informed consent from individuals before recording them or utilizing their data. Being transparent about the purpose of recording and how their information will be used is essential, empowering individuals to make informed decisions about their participation.

Our team will also implement robust data anonymization and minimization techniques to protect individuals' identities and limit the collection and storage of unnecessary Personally Identifiable Information (PII). By anonymizing data and minimizing the scope of collected information to what is strictly necessary, the risk of privacy breaches and misuse of sensitive information can be significantly reduced.

In considering the broader social context, it is important that we recognize that the impact of the AI Smart Mirror extends beyond the primary users to various stakeholders in society, such as the user’s support networks and traditional providers of fashion and wardrobe management services.

Because this product will encourage independence in individuals by providing accurate body measurements and virtual try-on capabilities, their caregivers and personal stylists could benefit from the adoption of our product. Caregivers would not need to focus as heavily on assisting with outfit selection, allowing them to improve other aspects of their support and engender more positive relationships with their dependents. Conversely, traditional providers of fashion advisory services may face challenges due to the disruptive nature of this technology, potentially leading to industry upheaval if its use becomes widespread, which could translate to job losses within the sector.

It is also important to recognize that introducing the AI Smart Mirror could intensify existing societal disparities while empowering individuals who can access and use the technology. Individuals who do not have access to or are unable to utilize the technology, such as those in lower-income communities or technologically illiterate populations, could be inadvertently excluded from benefiting from our product. Without appropriate rectification, this could perpetuate existing informational barriers and aggravate socio-economic inequalities. As such, once we begin large-scale production, we will look into ensuring equitable product access through potential partnerships with organizations working to improve technology literacy and working in lower-income communities.

By addressing these ethical considerations, we aim to build a product that not only provides significant value to its users but also upholds the highest standards of privacy, security, and social responsibility.

5. Hardware diagrams

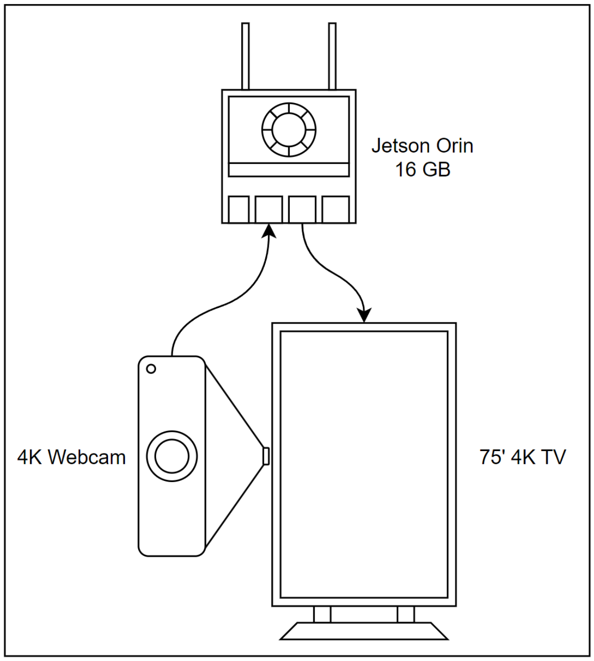

Figure 2: Hardware Block Diagram

The hardware architecture of the AI Smart Mirror is straightforward. There are three main hardware components for the client application:

75-Inch TV Screen: Functions as a mirror and display, providing a large and clear reflection of the user.

Jetson Orin: A powerful processing unit running Linux, responsible for image processing and running AI algorithms.

4K Camera: Captures high-resolution images necessary for accurate body measurement and outfit visualization.

The 4K camera, mounted on the TV screen, captures high-quality images of the user. These images are then processed by the Jetson Orin, which is connected to the TV and camera. The Jetson Orin runs sophisticated AI algorithms to measure the user's body dimensions and provide a virtual try-on experience.

The 4K camera captures an image of the user standing in front of the mirror. Next, the image is processed by the Jetson Orin, running the MediaPipe Pose module to analyze body measurements. Then, the processed measurements are used to display a selection of outfits on the TV screen. The user selects an outfit from the options displayed on the interface. The selected outfit is virtually applied to the user's image using the OOTD model with Stable Diffusion. The final image, showing the user wearing the selected outfit, is displayed on the TV screen.

The Jetson Orin sends video data to the TV over DisplayPort, and the camera connects to the Jetson Orin using USB.

6. Software diagrams and descriptions

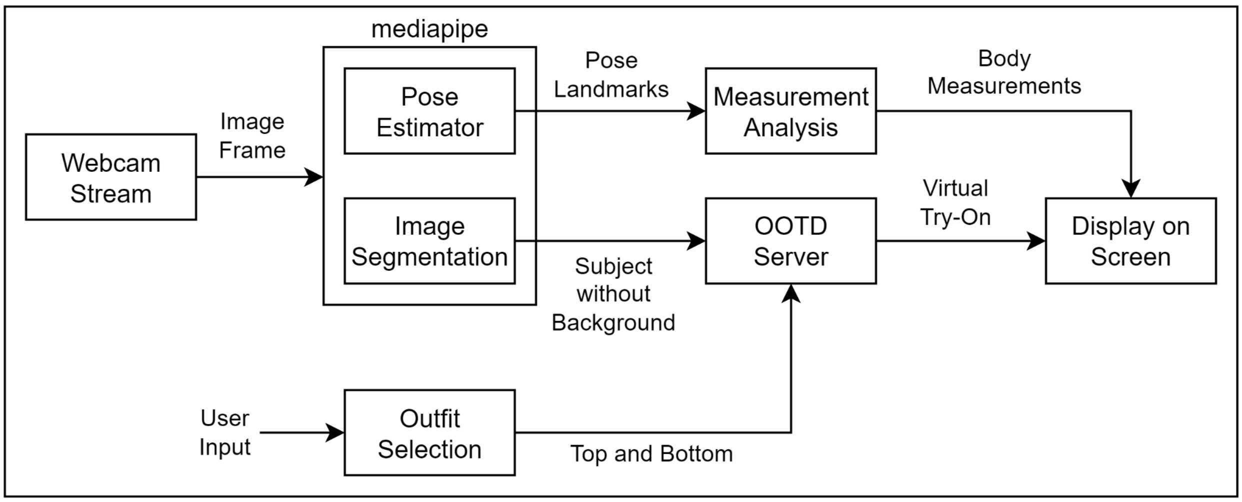

Figure 3: Software Block Diagram

The software architecture of the AI Smart Mirror is complex, involving many different libraries and even custom scripts to perform all the computation that is required for a robust and effective implementation of virtual try-on.

Webcam Input

We set up the webcam as a normal camera within the Linux OS. This allows us to access the webcam as a simple Linux device. Then, we use the OpenCV2 Python library to read individual frames from the webcam. We also perform rotate and mirror operations on the frame using OpenCV2 so that the image displayed on the TV looks like a mirror.

MediaPipe

MediaPipe is the image processing framework that we use to run both the pose estimation model and the image segmentation model. The pose estimation model returns the location of several body landmarks in screen space coordinates, which we will use to estimate body measurements on the user. The image segmentation model returns multiple masks for different object categories within the image. We use these masks to remove the background from the image to isolate the user. This improves the performance of the OOTDiffusion model later in the pipeline.

Body Measurement Analysis

The body measurement analysis is performed using completely custom code that was based on a combination of mathematical simplification and calibration.

For height measurements, we performed calibration to obtain a function that maps screen space coordinates to real world coordinates. We made the assumption that the user will always be centered on the screen at a constant distance from the screen. We then recorded 6 mappings from a known real-world height to screen space coordinates. Plotting these 6 points in a scatter plot revealed that the relationship between screen space height and real-world height is largely linear, so we used these points to derive a linear relationship that takes in a pixel location and returns the height of the user.

For shoulder and waist measurements, we first calculate the distance between the shoulder landmarks and the hip landmarks. These distances are based on the shoulder and hip joint of the user, so these distances do not reflect well the actual shoulder and hip measurements. So, we empirically determine scaling factors for both the shoulder and hip measurements to map joint-distance to real-world measured distance.

Outfit Selection GUI

The GUI running on the client application allows the user to select what top and bottom they are interested in trying on. The GUI also shows the user what options are available to try on. Once the user selects the clothing, the GUI starts the measurement process, which involves enabling video feedback for the user to position themselves on the screen. After 15 seconds, the client application takes the final picture of the user and starts to perform all the image processing. Once the server finishes generating the virtual try-on image, the GUI shows the user how they will look if they wore the clothes that they chose.

OOTDiffusion

The OOTDiffusion virtual try-on software gives us the ability to realistically put any clothing onto any user. There are many challenges with creating virtual try-on solutions. First, the target image is highly heterogeneous. Users can have many different sizes, shapes, colors, and poses. The image can also have many different lighting conditions and backgrounds. That being said, a simple overlay of the clothes on the source image will not look realistic and will be of limited use for the user.

Another approach is to use Stable Diffusion image generation models to “imagine” the user wearing the clothes. Since the image is generated from scratch, the image generation model can take into account differences in user body, pose, and lighting. This theoretically allows for much more realistic virtual try-on capabilities, although hallucinations and face preservation become new challenges with this approach.

For this project, we used OOTDiffusion which is an open-source diffusion-based virtual try-on project. We run OOTDiffusion on a remote server with much better computing capabilities than the Jetson Orin. This helps us reduce task latency and deliver faster results for the end user.

7. Quantitative requirements

7.1. Design Requirements

7.1.1. Camera Module:

- Camera resolution needs to be 4K to capture as much detail of the user as possible.

7.1.2. Microcontroller:

- Microcontroller needs to have two or more CPU cores to be able to run demanding vision algorithms on top of the base Linux OS

7.1.3. TV:

- TV resolution also needs to be 4K. At 75 inches diagonal, the TV is very large, so the resolution must be high to preserve visual integrity at very close distances.

7.1.4. OOTDiffusion Latency:

- The time taken for OOTDiffusion to finish generating the image should be less than 90 seconds. Anything more than 90 seconds would be considered uncomfortably long for the user.

8. Physical aspects

As mentioned above, the hardware architecture of the AI Smart Mirror is simple. Instead of repeating the hardware descriptions here, we included some pictures to provide an understanding of the hardware components.

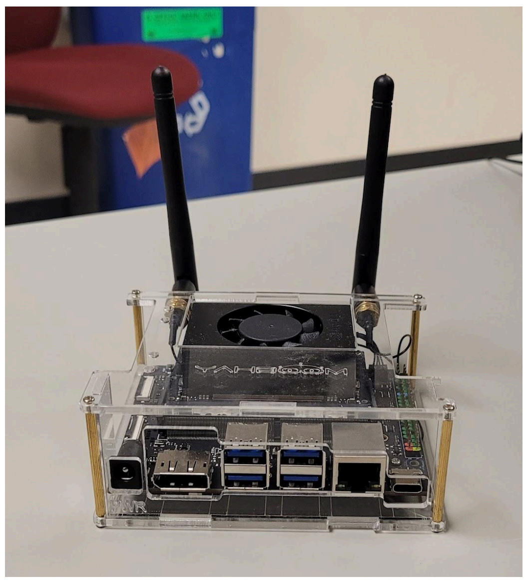

Figure 4: Jetson Orin Embedded Computer

Figure 5: 4K Camera

Figure 5: 75’ 4K TV

9. Standards

The following standards ensure the product is safe, effective, user-friendly, and compliant with regulatory requirements. It contains physical, coding and software, communication protocols, data privacy and security, and accessibility standards.

9.1. Physical Standards

9.1.1. ISO 9241-11: Ergonomics of human-system interaction - Ensures the usability and user comfort of the smart mirror interface.

9.2. Coding and Development Standards

9.2.1. ISO/IEC 25010:2011 for software quality requirements and evaluation (SQuaRE), ensuring the software within the glasses and associated applications meets high standards of functionality, reliability, usability, efficiency, maintainability, and portability.

9.3. Accessibility Standards

9.3.1. ISO 9241-171: Ergonomics of human-system interaction - Guidance on software accessibility: Provides additional guidelines to ensure that the software components of the AI Smart Mirror are accessible to all users, including those with disabilities.

9.3.2. Web Content Accessibility Guidelines (WCAG) 2.1: Ensures that the user interface is accessible to individuals with disabilities, including those with visual, auditory, and motor impairments. This standard guides the design of the interface to be inclusive and usable by a diverse user base.

9.4. Data Privacy and Security Standards

9.4.1. ISO/IEC 27001 for information security management, ensuring the protection of personal and sensitive data through risk management processes.

9.4.2. ISO/IEC 30141: This International Standard provides a reference architecture for IoT, including trustworthiness (security, privacy, and reliability) considerations, to guide the design and development of IoT systems.

9.5. Encryption Protocol

9.5.1. AES (Advanced Encryption Standard): For encrypting data at rest and in transit, AES will be used. It is widely used for its robustness and efficiency. It's crucial for protecting sensitive information, including PII and biometric data.

9.5.2. TLS (Transport Layer Security): TLS is essential for secure communication over a network. Implementing TLS 1.2 or higher for device communications can safeguard against eavesdropping and tampering.

10. Gantt chart

A Gantt chart Including a list of all tasks, distribution by individuals, and estimated timeline can be found here.

11. Parts

Here's a breakdown of the components used for the smart mirror and the estimated costs based on common market prices for similar items. Moreover, the link for each part is attached to the item name.

11.1. Main Parts

11.1.1. Yahboom Jetson Orin NX 16GB

- Cost: $1029.99
- Description: An embedded computer from Nvidia with a dedicated GPU and AI hardware.

11.1.2. Sony Bravia EZ20L 75'' 4K Monitor

- Cost: $1350.00
- Description: A large-format TV to be used as the mirror.

11.1.3. VIVO Mobile TV Cart

- Cost: $99.99
- Description: A wheeled cart for the TV to allow for easy transport of the TV.

11.1.4. Spedal 4K Webcam

- Cost: $89.99
- Description: High resolution and wide angle webcam to capture the image of the user.

Component	Quantity	Unit Cost	Total Cost
Yahboom Jetson Orin NX 16GB	1	$1029.99	$1029.99
Sony Bravia EZ20L 75'' 4K Monitor	1	$1350.00	$1350.00
VIVO Mobile TV Cart	1	$99.99	$99.99
Spedal 4K Webcam	1	$89.99	$89.99
	Total Prototype Cost:		**$2569.97**

The total cost of this project is very high. This amount also does not take into account the cost of using cloud computing for OOTDiffusion. As it stands, this price point is only accessible for large-scale retailers, limiting the total market for this product.