For use by the Project lecturer	Approved		Revision required	

Feedback

All sections changed to use template fonts.

To be completed by the student				
PROJECT PROPOSAL 2022	Project no	HG9	Revision no	1
Title Mr	Surname **Williams**	Initials **ML**	Student no 18013555	Study leader (title, initials, surname) Mr. H. Grobler
Project title **Real-time hand gesture control of a virtual object in augmented reality**				

Language editor name	Language editor signature
Mitchell Williams	
Student declaration I understand what plagiarism is and that I have to complete my project on my own.	**Study leader declaration** This is a clear and unambiguous description of what is required in this project. Approved for submission (Yes/No)
Student signature	Study leader signature and date 2022-06-20

1. Project description

What is your project about? What does your system have to do? What is the problem to be solved?

Augmented reality is a powerful tool to interact with computer interfaces in a natural environment using intuitive human gestures. This project aims to implement a real-time system that can recognize the gestures and positions of a human user's hand and interpret them as different input commands to a virtual object that is instantiated in a live video feed. The system has to recognize the hand gestures and match them against a collection of known gestures and associated virtual object commands. The system must apply the relevant commands to the virtual object in real time with no discernable delay to the user. The virtual object can be handled by the user - moved in multiple directions and rotated in the context of the environment surrounding the user. The virtual object must also interact with the environment it is projected into so that it does not merely float against the background but instead rests on a surface and cannot be pushed through objects in the video stream - like a real 3D object would not be able to be pushed through a solid object.

Made clear that system will be attempted to be implemented on an embedded platform -additional notes on expectations regarding system performance on embedded system VS PC platform implementation contained in lab book.

2. Technical challenges in this project

Describe the technical challenges that are *beyond* those encountered up to the end of third year and in other final year modules.

2.1 Primary *design* challenges

The design of the image processing and hand recognition algorithms will be the principle design challenge of this project in addition to the 3D object generation and image theory required to integrate a virtual object with objects present in a real image. First principles design of a system capable of recognizing the different gestures and positions a hand can make is a primary challenge due to the difficulty of emulating the human visual recognition system. Designing algorithms to integrate a 3D object into a video feed will be another key challenge because of the complexity of three-dimensional environments. The complex nature of video processing, computer vision and three-dimensional scene reconstruction of the environment are design challenges beyond the scope of previous years' work.

Implementation on an embedded platform will be the principle challenge of the project in addition to proof-of-concept system performance on a PC

2.2 Primary *implementation* challenges

Implementing the hand recognition as well as virtual object manipulation algorithms in real-time on an embedded platform with no discernable delay to the user will be the key implementation challenge. Implementing this system in a real-world setting with the visual occlusions, noise and unpredictable lighting conditions present in a standard office will be a challenge. Inserting even a rudimentary graphical object into the context of a real environment from a video feed will be an implementation challenge due to the unpredictability of those environments and doing this while contending with processing constraints and video transfer speed limitations will be a difficult implementation challenge on the embedded system.

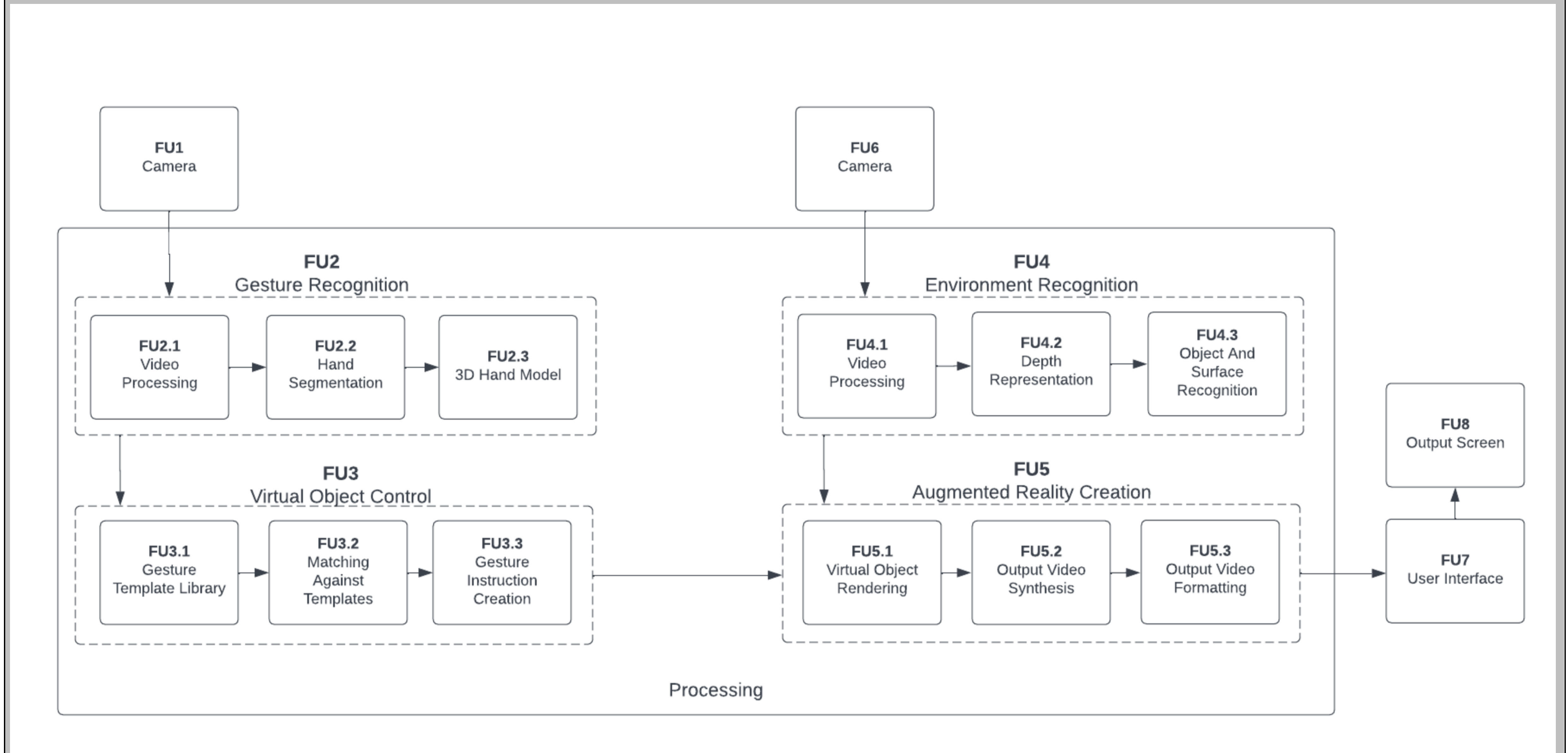
3. Functional analysis

3.1 Functional description

Describe the design in terms of system functions as shown on the functional block diagram in section 3.2. This description should be in *narrative format*.

The user positions themselves in front of a camera (FU1) which captures continuous video of the user and their upheld hand as input to the gesture recognition algorithm (FU2). The video input is first formatted and its individual frames extracted (FU2.1) for the correct input to further processing. Image segmentation (FU2.2) is performed to distinguish the user's hand from the rest of the image. This segmented image is then converted into a 3D model of the user's hand (FU2.3) which is the input to the virtual object control algorithm (FU3). This algorithm first accesses a template library (FU3.1) of existing hand models for different input gestures and then matches the current input hand model to any similar templates (FU3.2) stored in the library - identifying the gesture. This identified gesture is then combined with the parameters of the input model (FU3.3) and sent as a gesture instruction to the augmented reality creation algorithm (FU5). At the same time, another camera (FU6) captures continuous video of the environment in front of the user and inputs it to the environment recognition algorithm (FU4). Here the input video is formatted and has its individual frames extracted (FU4.1) in order to create a depth representation of the environment around the user (FU4.2). This depth information is then used to recognize objects, boundaries and surfaces (FU4.3) that the virtual object can interact with or be affected by. This information is then used in the augmented reality creation algorithm (FU5) where a virtual object is rendered using graphical methods (FU5.1) and based on the gesture and environment input, the two video streams are meshed together (FU5.2) to create a cohesive output video which is formatted (FU5.3) for output to the user interface (FU7). The user interface shows the rendered virtual object in the environment, the current model of the user's hand as well as any error or log messages. The user interface is displayed to the user on a screen (FU8).

3.2 Functional block diagram



4. System requirements and specifications

These are the core requirements of the system or product (the mission-critical requirements) in table format IN ORDER OF IMPORTANCE. Requirement 1 is the most fundamental requirement.

	Requirement 1: the fundamental functional and performance requirement of your project	**Requirement 2**	**Requirement 3**
1. Core mission requirements of the system or product. Focus on requirements that are core to solving the engineering problem. These will reflect the solution to the problem.	The system must allow a user to control and manipulate a virtual version of a simple three-dimensional geometric shape object in augmented reality using hand gesture control in real-time.	The system must be able to recognize nine of a user's discrete hand gestures from video input of a single one of the user's hands in real time.	The system must be able to create, render and manipulate a virtual object in a manner consistent with how a real-world version of the same object would be able to be be manipulated and controlled.
2. What is the target specification (in *measurable* terms) to be met in order to achieve this requirement?	A virtual 3D cube that is integrated into the video feed of the environment must be able to be manipulated by a user's hand gestures as well as interact realistically with the environment all with a latency of less than 41.6ms. (24fps)	The system must be able to correctly identify the 9 discrete static hand gestures needed for interacting with the virtual object with a latency of less than 41.6ms from a camera input. Made clear that 41.6ms latency comes from frame rate of 24fps. 1/24 = 41.6ms	A 20cm-sided three-dimensional cube must be rendered, and then the user must be able to rotate it 360 degrees in the x, y and z-plane, move it up and down, backwards and forwards as well as side-to-side by 10cm with gesture input.
3. Motivation: *how* or *why* will meeting the specification given in point 2 above *solve the problem*? (Motivate the *specific* target specification selected)	The illusion of augmented reality is created if the object is controlled and rendered at a frame rate of 24 fps - the lowest refresh rate possible before the human eye and brain begins to notice lag and not perceive fluid motion.	These gestures represent the up, down, left, right, backwards, forwards and rotate gestures and being able to identify them in less than 41.6ms allows the system to recognize a user's intended gesture control in apparent real-time (24fps).	These actions represent all the ways an object can be moved in the real-world and thus a cube that can be moved in a virtual setting in these directions and orientations behaves similarly to a real-world cube and creates "augmented reality."
4. How will you demonstrate at the examination that this requirement (point 1 above) and specification (point 2 above) has been met?	Hand gestures will be performed in order to manipulate the object in a desired manner and the corresponding virtual object movement as well as environment's response will be demonstrated as well as the latency taken to perform this action.	A live representation of the user's hand will be displayed in the user interface and 9 gestures will be performed and their correctly interpreted gestures demonstrated, along with the time it took for the system to interpret the gesture.	The cube will be rotated 360 degrees in each planar direction, and then moved sequentially 10cm in every direction on the surface it is resting on to demonstrate the manipulation and rendering of the virtual object.
5. Your own design contribution: what are the aspects that *you will design and implement yourself* to meet the requirement in point 2? If none, *remove this requirement.*	The virtual object control algorithm, hand segmentation, hand model, gesture interpretation algorithm, user interface, environmental depth representation, object collision avoidance and surface detection will be designed from first principles.	The hand segmentation and feature extraction algorithms as well as construction of the hand model will be implemented from first principles.	The virtual object manipulation algorithm which includes scaling, translating, placing and rotating the object based on gesture input will be implemented from first principles.
6. What are the aspects to be taken off the shelf to meet this requirement? If none, indicate "none"	Input cameras, an embedded platform, display, image capture, image to array conversion and image display libraries as well as the graphical rendering of the virtual object will be taken off-the-shelf.	A camera, embedded platform and display will be taken off-the-shelf. Additionally, image processing libraries to format and convert the video frames to pixel arrays will be taken off-the-shelf.	The graphical rendering of the virtual object will be taken off-the-shelf with a library.

Cinematography industry standard of 24fps refresh rate is chosen due to it being the lowest refresh rate a video feed can be updated while appearing to show fluid motion to the human eye and brain and due to processing constraints - higher would require more processing

Made clear that the embedded platform will be taken off the shelf and that image processing libraries will only be used for low-level operations such as capturing an image from a webcam, displaying it on a screen and converting the image to an array of pixels for further computations.

System requirements and specifications page 2

	Requirement 4	Requirement 5	Requirement 6
1. **Core mission requirements of the system or product.** Focus on requirements that are core to solving the engineering problem. These will reflect the solution to the problem.	The system must function in user-apparent real time with no visible latency to the user.	The virtual object must interact with the real environment it is rendered inside of in a physically realistic and consistent manner.	Requirement 6 removed due to redundancy - its function had already been implied by previous requirements.
2. **What is the target specification** (in *measurable* terms) to be met in order to achieve this requirement?	The virtual object's position and orientation as well as the model of the user's hand must be updated 24 times a second - creating a 24fps video.	The virtual object must remain static for more than 10s on any flat surface in the environment if no input gestures are provided to it and only be moveable 10cm around or over objects in the environment - not through them.	
3. **Motivation**: *how* or *why* will meeting the specification given in point 2 above *solve the problem*? (Motivate the *specific* target specification selected)	Human eyes interpret a video that has a frame rate of less than 24fps as choppy and disjointed thus refreshing the virtual object and model of the hands at 24fps creates a smooth and seemingly real-time representation of these objects.	By preventing the virtual object from moving through other objects and surfaces the sense of realism is upheld and the virtual object appears to behave the same way an object in the real world would.	
4. **How will you demonstrate at the examination** that this requirement (point 1 above) and specification (point 2 above) has been met?	A frame counter will be implemented in the user interface to display how many times per second the virtual object and model of the hand are updated - if the number the frame counter displays is 24 or higher the system will appear to run in real time.	The virtual object will be rendered onto a flat table and a physical 20cm-sided cube will be placed onto the table too. The virtual object will be attempted to be moved through the real cube and the resulting behaviour and distances demonstrated.	
5. **Your own design contribution:** what are the aspects that *you will design and implement yourself* to meet the requirement in point 2? If none, *remove this requirement.*	The frame counter, user interface, model of the hand and virtual object manipulation algorithm will be implemented from first principles.	The creation of a depth-representation of the environment, collision avoidance as well as boundary detection algorithms will be implemented from first principles in addition to the virtual object manipulation algorithm.	
6. **What are the aspects to be taken off the shelf** to meet this requirement? If none, indicate "none"	The graphical rendering of the virtual object as well as input cameras will be taken off-the-shelf.	An input camera, embedded platform and display will be taken off-the-shelf. Off-the-shelf nature of embedded platform mentioned again where necessary.	

5. Field conditions

These are the REAL WORLD CONDITIONS under which your project has to work and has to be demonstrated.

	Field condition 1	Field condition 2	Field condition 3
Field condition requirement. In which field conditions does the system have to operate? Indicate the one, two or three most important field conditions.	The system must function in standard indoor lighting conditions for computer work	The virtual object will be manipulated using a single visible hand at a time.	The virtual object must be able to be manipulated when it is placed in the user's immediate vicinity.
Field condition specification. What is the specification (in measurable terms) for this field condition?	The room must be at 300lux brightness or greater.	A maximum of 1 hand at a time can be present in the camera feed and a maximum of 1 finger can be partially occluded by the hand it is attached to.	The virtual object must be placed within a 2m radius of the user and not more than 2m away from the camera.

6. Student tasks

6.1 Design and implementation tasks

List your primary design and implementation tasks in bullet list format (5-10 bullets). These are *not* product requirements, but *your* tasks.

Grammar of all tasks modified to not be in the imperative form.

- Research will be conducted on how gesture recognition is performed using hand models and the calculations involved.
- Two cameras and a depth-sensing device must be interfaced with a PC to enable communication and image transferral.
- Image processing algorithms for compressing and formatting video input must be implemented.
- A hand image segmentation algorithm must be designed and implemented to separate the pixels of a hand from the background of a camera image.
- A hand model creation algorithm must be designed to create an accurate model of a hand and its current gesture.
- A gesture template matching algorithm must be implemented to match an input gesture to an assembled template library of gestures.
- A virtual object rendering algorithm must be developed to create a virtual object and insert it correctly into a live video feed.
- It must be ensured that all algorithms and processing is performed in real-time with no discernable delay to the user.

6.2 New knowledge to be acquired

Describe what the theoretical foundation to the project is, and which new knowledge you will acquire (*beyond* that covered in any other undergraduate modules).

- Understanding of computer vision techniques such as image segmentation and feature extraction will be developed.
- Knowledge of statistical methods for confirming the accuracy of machine learning models will be developed.
- The techniques of artificial intelligence and deep learning will be researched and used to classify hand image data.
- The background of real-time video compression, image processing and analysis will be required in order to handle the large amounts of input video data required.
- An understanding of computer graphics, visual interfaces and three-dimensional digital representations of objects will be developed.
- The mathematics behind computer vision and image processing will be researched and utilized in order to synthesis the output video.