Daniel Eisen (Team 3)
300447549

ECEN 425 - Assignment 3 - Technical Report

Form Tracking - Formfit

1. Introduction

This team's initial design pitch was focused on the delivery of a fully-integrated home gym experience (the *CableSense*) with a focus on smart features such as realtime form feedback and spotting. However, as we went into the market research phase and received feedback from potential users we decided that our main target problem should be form tracking and feedback. As such we now present the *Formfit,* a central camera hub with modular expandability and remote feedback/interfacing. This report will then discuss our solution to tracking the users position; the technology and software, and the required hardware and its constraints.

2. Form Tracking Approach

The *Formfit's* main feature will be a form tracking system responsible for estimating the user's body position during exercises. This will include the cameras, pose estimation software, and wearable tracking tech. This feature will be available for loaded preset exercise routines and provide immediate corrective feedback to the user about stance and form to better their exercise (outlined further in Wills report), as well as collecting performance data for progress summary and tracking.

Keystone points of the user will be tracked via shipped wearables with markers of known, either velcro style compression wear or attachable patches.

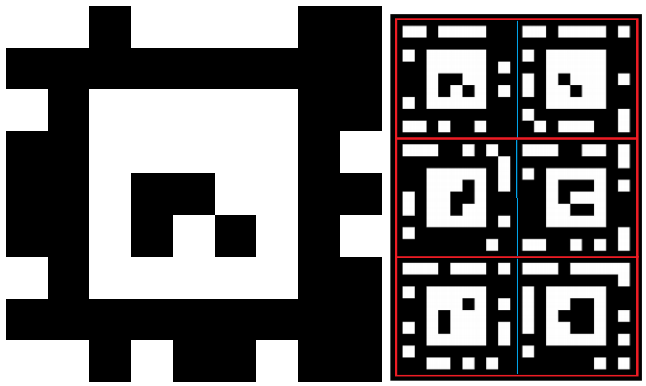
From this the pipeline can be basically outlined:

- Image sensor captures frames
- Hardware codec compresses video
- Ingest frame for processing via openCV
- AprilTag API detects trackers
- Frame space coords converted to volume space
- With Keystone points and error from proper form sent for user feedback

Daniel Eisen (Team 3)
300447549

3. April Tag Tracking

AprilTags are a form of **fiducial marker,** or more concisely a **reference marker** that computer vision software uses to perform specific operations based on the marker's type, location, and any encoded data. For example deriving that marker's position within a 3D volume. There are various types of these markers but this report will discuss the usage of AprilTags.



(a) (b)

Figure 1: AprilTag individual example (a), group of UID'd tags (b)

April Tags are a specific type of fiducial marker that's software (generation and detection api) is openly licensed (BSD) [1]. As seen in *figure 1a* there are high contrast, relatively simply patterned tags that can be compared to low data size QR code. That contrast and their intentional lack of rotational symmetry allows the detection API to quickly detect, isolate and orient the tags. The small amount of encoded data is the unique ID of the given tag, see variants in figure 1b, and this is what will be used to identify where on the body this marker is placed.

3.1 Software Pipeline

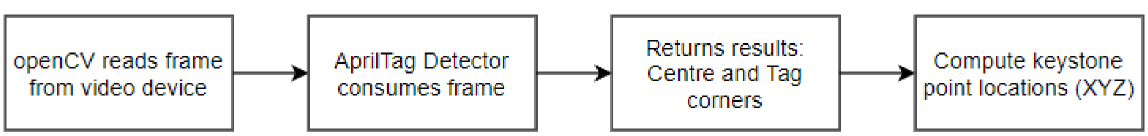
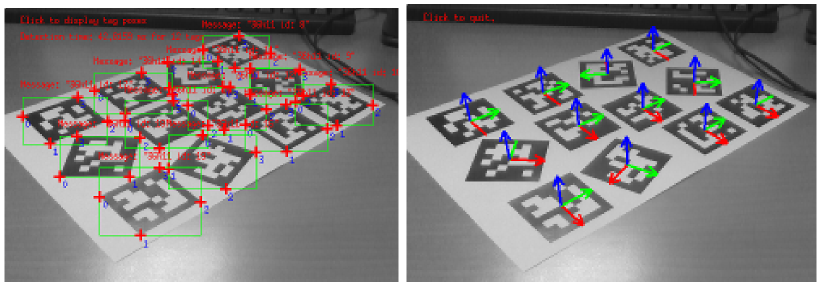


Figure 2: Tag detection and tracking

As this platform will be running within a linux environment (for the network stack etc) this application will have the openCV/openMV and dependent AprilTag API libraries available for use. Therefore the process can be abstracted into the following tracking pipeline:



(b) (b)

Figure 3: Tag detection and pose estimation

1. OpenCV `VideoCapture` object setup
    a. Frame is `read` from capture stream
    b. Frame is converted to grayscale (if in color)
2. AprilTag detector is initialized with tag family ID (which flavour of tag is specifically in use)
    a. `Detector` consumes frame
    b. `Results` are returned containing a list of tag centers and of corner XY coords and IDs for every tag detected with the frame, *figure 3a.*
3. From known camera focal length and tag dimensions the pixel information is used to derive the scale of the frame relative to the volume. This derivation is done by the API using these as input variables.
4. From this the API (specifically `PoseEstimationMethod` or `solvePnP`) can compute the pose of each tag ID within the frame, *figure 3b.*
    a. XYZ positional data
    b. Euler tag rotation

Using this approach it is feasible to achieve sub pixel positional accuracy.
As other reports outline, during development, a dataset of ideal form will be generated with the assistance of hired consultants. These poses are then passed as input to the comparative NN to generate an error from the ‘ideal’ form for a given exercise that will then drive the user feedback and performance tracking outlined by Will.

3.2 Proof of concept

As a proof of concept test, 3D printed wearables wear made and a desktop version of the tag pose tracking see figure 4. This was done from a single perspective USB webcam at 30fps and proved the feasibility of the approach as the markers could be tracked in real time and maintained continuous pose data per ID.

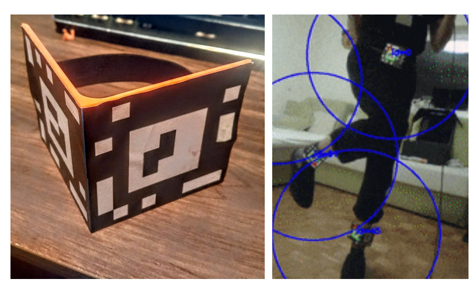


Figure 4: Test tag and tracked points

Daniel Eisen (Team 3)
300447549

4. Camera

In choosing suitable camera modules to be the driving sensor for the units a few specifications must be met/considered to maintain tracking functionality, not saturate the devices bandwidth, and support a universal interface for data output to enable easy module upgrade/changes.

4.1 Image Specifications

Full-colour out is not necessary due to the contrast based fiducial marker system that only processes brightness data. The expected volume of use is ‘small room scale’ i.e. 1-5 metre square. With this smallers use scale motion becomes more important for effective use so a sustained fps at lower resolution is preferred.

From this spec, and the above testing a minimum image output of 1280x720@30fps, a fixed focal spec of greater than 0.5m (i.e in focus from 0.5m outward) and any FOV over 70 to ease positional requirements.

4.2 Compression Processing

In order to enable the wireless aspect of our modular design the captured image data must occupy a compressed bandwidth. Wifi-Direct supports a max of 250Mbps. As such the selected camera module would require an onboard image processor for real-time hardware compression, such as a JPEG engine or H.264 codec.

To validate a minimum spec or 720p@30fps a medium quality motion-jpeg (MJPG) encoded video stream would occupy 27-30Mbps and h.264 at 2Mbps with the caveat of less hardware availability.

4.3 Suitable Solutions

OmniVision supplies a range of camera/processor module lines that meet these specifications, with output available in USB, DCMI, CSI.
Two of which can be deemed sufficient for this design:

OV9726: USB/DCMI output, 1MP max, 720p@30fps with codec MJPEG
OV5647: CSI/DCMI output, 5MP max, 1080P@30fps with codec H.264

Either of these modules (or any within their line) will provide enough data for smooth tracking, provide the necessary compression to meet bandwidth restrictions

Daniel Eisen (Team 3)
300447549

References

[1] “AprilTags Visual Fiducial System” https://april.eecs.umich.edu/software/apriltag

#Qnty	**Cost (ex GST) / each**
OV9726	NZD 5.21
OV5647	NZD 2.64
Estimated wearable wholesale cost	Assume bulk purchase from eg china: 3-10 NZD manufacturing cost per item 15 NZD international shipping *ref: sewport 2020*