EN 601.456 Computer Integrated Surgery II Project Proposal

# Vital Monitor and ID Detection through Machine Vision for Improving EMS Communication Efficiency

Robert Huang
Biomedical Engineering Senior
rhuang22@jhu.edu

February 2021

## Contents

# 1 Abstract

Glasses that provide live feed are currently used by emergency medical technicians simply as a camera. Consequently, there is opportunity to use the live feed, along with AI, to streamline certain processes. In particular, this project seeks to use this live feed to: 1) allow remote doctors to gain visual access to the information available from medical devices (such as ultrasound) in order to facilitate physician guidance, and 2) automatically pull key information from standard documents like driver's licenses (by holding up identification to the glasses) to fill in personal data in forms and databases so that medics may spend less time on paperwork and more time on treating patients. The first objective's basis is to use deep learning, along with computer vision, to detect and deform medical data screens such that a person may remotely view them head-on. The second objective's basis is to use deep learning to detect IDs, and then to use optical character recognition and generic text parsing to extract the personal information. Training sets include those of desired monitors, as well as those of desired IDs. By the end of the project, at minimum, this work hopes to develop the ID detection and extraction algorithm, and at maximum, this work hope to have built and incorporated these algorithms into the workflow of a current model of smart glasses.

# 2 Clinical Motivation

Smart glasses, while themselves are not a recent invention, are only just being implemented into healthcare environments. These glasses are most often used by surgeons or emergency medical workers to record and live-stream video feed from, respectively, surgical rooms and on-the-field operations [1] [4]. This feed is fed through a smart phone and then sent over the internet to remote healthcare workers, who can either provide their advice and assistance for the current patient, or, if they are trainees, learn from watching these healthcare procedures [1]. Currently, only the raw live footage is used. Therefore, there is opportunity in taking this feed and running it through artificial intelligence in order to streamline healthcare processes to ultimately increase efficiency of patient care. In particular, there are two objectives:

1. **Provide View of Data Monitors Remotely with AI and Computer Vision**

   During an emergency response on-the-field, it is often necessary to obtain advice, assistance, and/or clearance from a remote physician to carry out a procedure. To improve care, the remote physician should have direct access to the information available from medical devices used by the medics, especially vitals signs monitors and ultrasound scanners. With visual confirmation of vitals, physicians are 2-3x more confident and give treatments 2-3x faster [4]. Furthermore, without clear view and recording of vitals, 43.4% of information can be lost, resulting in repeated procedures once the patient reaches the hospital [4]. Therefore, it is pertinent to obtain the visuals of these data monitors. Unfortunately, while certain data monitors that come packaged with smart glasses will provide direct access of the data to physicians, all in all there are poor interoperability standards [3]. However, the human-readable displays of these devices are always readable when in view of the smart glasses camera. Therefore, this project aims to take the video feed of these devices and incorporate machine vision to provide physicians direct visual access to data monitors on the field.

2. **Automatically record information into a digital medical note from IDs**

   During a typical emergency medical response, 25-65% of the patient's care is spent documenting [5], with 1-10 minutes of it being spent on obtaining and recording 'simple' information of both the patient and health care workers, such as names, birth dates, ages, addresses, or ID numbers [2]. Therefore, streamlining patient identification by pulling key information from standard documents like driver's licenses means that medics spend less time on paperwork and more time treating patients. Consequently, this project seeks to produce a deep learning and optical character recognition (OCR) algorithm that detects identification, extracts the relevant information, and writes directly into the necessary documents, such as an ePCR.

## 3 Goals and Significance

As noted in the previous section, current smart glasses are used solely for their raw video and audio feed. It is believed that this feed can be analyzed through artificial intelligence to streamline processes in healthcare. Therefore, the goal of this project is to develop the elements of artificial intelligence that will take the video feed and perform two objectives. The main objectives in this project will be to provide direct visuals of on-the-field data monitors to remote health care workers and to extract and input information from identification into digital medical notes. Providing visuals will improve on-the-field outcomes by increasing confidence and treatment speeds, and by reducing repeated procedures. Extracting information from identification will improve on-the-field outcomes by moving time spent on documentation to patient care. Furthermore, this project's success will serve as an initial assessment into the viability of artificial intelligence in smart glasses in a health care environment.

# 4 Technical Approach

## 4.1 Overall Workflow

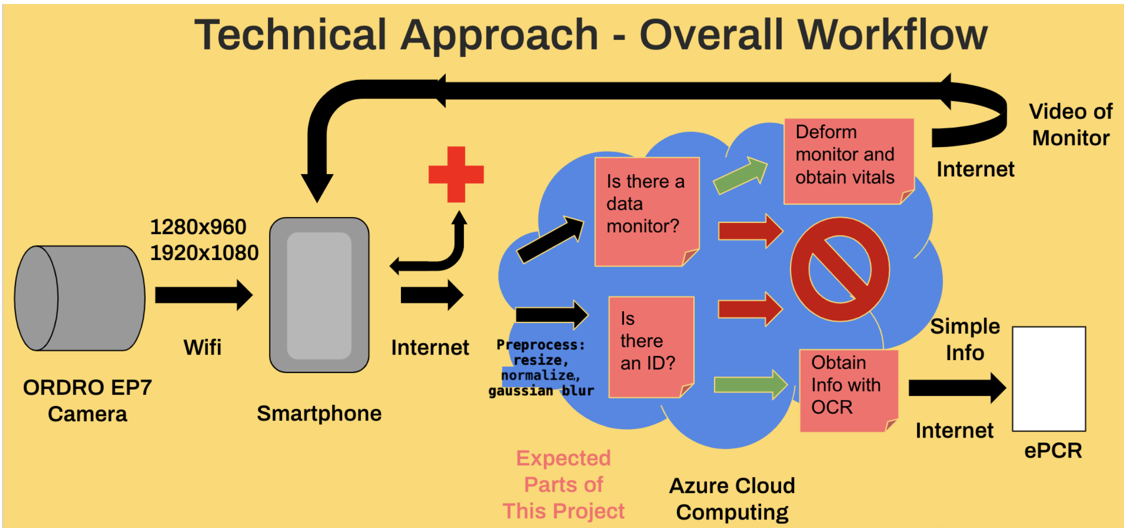


Figure 1: Overall Ideal Workflow from Smart Glasses Camera.

The ideal overall workflow of the smart glasses, in relation to this project, is shown in Figure 1. Video feed of images of size 1280x960 and 1920x1080 will be sent to the smartphone through Wifi. From there, these images will be sent to the remote healthcare provider and over the internet into the cloud, where the algorithm will process the video. Prior to reaching the algorithms, the video feed will be preprocessed through resizing, normalization, and gaussian blur to ensure standardization and removal of noise. For either algorithms, it will first check if its respective object is within the image. If the object is detected, the algorithms will process the images respectively. If a data monitor is detected, the pixels of the monitor will be cropped out, its features will be detected and registered, and the image will be deformed such that it appears head-on. This image will be sent back to the smartphone, and the image will be appended to the video feed and sent to the remote health care worker. If identification is detected, the algorithm will grab text from the image, parse the text, and send the information over the internet to an electronic medical note, such as an ePCR.

## 4.2 Dataset

- Dataset IDs: 200 800x800x3 images of driver licenses from all 50 states, with varying amounts from a particular state.
- Dataset Medical Devices: 60 800x800x3 images of Medical Devices with Monitors. There are three main medical devices with 20 images of different views of them: Butterfly Digital Ultrasound, Lifepack 15 Vitals monitor/defib, Zoll X series monitor/defib.
- Dataset Characters: 20x20x3 images within CHARS74k: 64 classes (0-9, A-Z, a-z), 7705 characters obtained from natural images, 3410 hand drawn characters using a tablet PC, 62992 synthesised characters from computer fonts.

## 4.3 Objective 1:Detection and Deformation of Data Monitors

There are three main steps in order to obtain a head-on image of the data monitor.

1. **Detect the display of a vital signs monitor or ultrasound image.**
   In order to detect the presence of a monitor, a YOLO Deep Learning Framework will be constructed and trained on the medical devices dataset. YOLO was chosen due to its aptitude in processing large data at many frames per second. Furthermore, according to the problem, it is only necessary to detect whether one object (which monitor?) is present in the frame, which YOLO excels at.

2. **Deform monitor such that the monitor appears head-on** As soon as a monitor is detected, it is necessary to deform it for good visualization of the vitals. In order to deform the monitor, it is necessary to first determine the current orientation of the monitor. Due to its scale and affine invariant properties, a "Blob Feature" registration algorithm will be written. Once the live monitor's features are detected, the features will be compared to a control image of the monitor and subsequently deformed to appear head-on.

3. **Read vitals signs data** Once the monitor is deformed, an OCR program that uses a state of the art deep learning technique by Wojna et al, 2017 will be run on the image to read vital signs.

For each code, the target accuracy will be 95%, with a decision speed of ¡8ms (to support a 60 frames per second camera). Ultimately, a combined accuracy of 95% with a decision speed of ¡16ms is desired.

## 4.4 Objective 2:Detection and Information Extraction of IDs

There are two main steps in order to detect and extract information from identification.

1. **Detect the presence of an ID**
   In order to detect the presence of an ID, the same YOLO Deep Learning Framework will be constructed and trained on the IDs dataset. YOLO was chosen due to its aptitude in processing large data at many frames per second. Furthermore, according to the problem, it is only necessary to detect whether one object is present in the frame, which YOLO excels at.

2. **Read and categorize information on the ID.** Once an ID is detected, the same OCR program used in Objective 1 will be run on the image to obtain text. Generic Text Parsing will then be used to sort the information into desired categories, such as name, birth date, etc.

For each code, the target accuracy will be 95%, with a decision speed of ¡8ms (to support a 60 frames per second camera). Ultimately, a combined accuracy of 95% with a decision speed of ¡16ms is desired.

# 5 Deliverables

The main deliverables and their respective expected deadlines for this project are listed in the table below. They are categorized as either a minimum, expected, or maximum activity/deliverable.

| | Activity | Deliverable | Expected Completion |
|---|---|---|---|
| **Min** | Obtain datasets of IDs and Medical Device with Monitors, and characters. | Dataset of IDs, Medical Devices, and characters | 2/19 |
| | Code the YOLO framework for detection of IDs. Train on ID dataset. Code the OCR framework. Train with character dataset. Code Generic Text Parsing Code and assess its performance with a constructed ground truth dataset. Combine code to read IDs. | Documentation of overall code and performances of each section. | 3/15 |
| **Expected** | Duplicate and Train YOLO on Medical Device dataset. | Documentation of code and performance of YOLO on Devices. | 3/20 |
| | Code Blob Feature Detection and deformation algorithm. Qualitatively assess how many deformations are acceptable. | Documentation of Blob Feature Algorithm and its performance. | 3/30 |
| **Max** | Incorporate the algorithms into the current overall workflow. Assess their performance in the workflow using a written test. | Documentation of incorporation. Testing procedures for performance assessment. Results of the tests. | 5/1 |

Figure 2: The minimum, expected, and maximum key activities with their corresponding deliverables

Overall, the minimum activity is to obtain datasets for identification, desired medical devices with monitors, and alphanumerical characters, as well as to fully develop, test, and document the algorithm to detect the presence of identification, extract text from the identification, and to categorize the text into desired information. The expected activity is to develop, test, and document the algorithm to detect the presence of a desired medical device and deform its display such that health care workers view it head-on. The maximum activity is to document and test the incorporation of both of the algorithms into the overall workflow for the smart glasses. Deliverables will be in the form of datasets, tables, graphs, videos, code, and documentation.

# 6 Dependencies

There are five primary dependencies that the team members do not have direct and personal control over. The dependencies are listed in the Figure 3 below, along with the rationale for their need, their current status, their contingency plan (should they not be obtained on time), and their planned and hard deadlines.

| Dependency | Need | Contingency | Status | Planned Deadline | Hard Deadline |
|---|---|---|---|---|---|
| ID/Medical Device/ Characters Datasets | For training and assessing algorithms | Begin coding without training. | Currently met | 2/19 | 3/20 |
| Computer/Internet | For coding and communication | Use public computers. Use mobile data. | Currently met | Continuous | Continuous |
| MDAirSupport Sample Product | For testing for incorporation into overall workflow | Write procedures for other MDAirSupport to test. | Am being promised one | 4/1 | 5/1 |
| Deep Learning Mentor | For Optimizing Algorithm to reduce computation | Spend time researching optimization | Currently met | Continuous | Continuous |
| Microsoft Azure or other Cloud Computing | For incorporation into overall workflow. May also require for training. | For training: use public JHU computer. | Under consideration: If needed, will email. | 4/1 | 5/1 |

Figure 3: The dependencies with their contingency, status, and deadlines.

The datasets that are required for training will be obtained from the internet (IDs and alphanumerical characters) and from MDAirSupport, who have the desired medical devices and have taken multi-view images of them. The computer/internet dependency will be obtained from the team's primary internet provider, or from the school. Both the MDAirSupport Sample Product, which includes the smart glasses, and the Cloud Computing Network will be obtained from MDAirSupport, who have promised their assistance once the algorithms are written and tested. Finally, the Deep Learning Mentor will be Dr. Mathias Unberath, from JHU's Computer Science Department.

# 7 Timeline

The table below lists the milestones and timeline for this project, including deliverables, class reports, and class presentations. Ideally, the project will proceed in a strictly linear manner. The only exception to this will be the dependencies, which will be resolved concurrently. Notably, the only dependencies that have yet to be obtained will be due on April 1st, when the "Maximum" deliverables are expected to proceed.

| Milestone | Deadline |
|---|---|
| Obtain datasets. | 2/19 |
| Code YOLO | 2/22 |
| Train YOLO on IDs and record results | 2/25 |
| Code OCR | 2/29 |
| Train OCR on characters and record results | 3/1 |
| Create Ground Truth Set for Generic Text Parsing | 3/5 |
| Code and Assess Generic Text Parsing and record results. | 3/15 |
| Combine all individual codes to read IDs | 3/15 |
| Duplicate YOLO and train on Devices Dataset and record results | 3/20 |
| Checkpoint Presentation | 3/23 |
| Generate Test Set for Blob Feature Deform Algorithm Testing | 3/25 |
| Code Blob Feature Deform Algorithm | 3/28 |
| Assess Blob Feature Deform Algorithm and record results. | 3/30 |
| Investigate how to write to ePCR. | 4/5 |
| Feed video back into smartphone. | 4/10 |
| Incorporate all algorithms into workflow. | 4/15 |
| Generate tests to assess the functionality of the application. | 4/20 |
| Paper Presentation | 4/22 |
| Assess overall application functionality. | 5/1 |
| Final Paper | 5/4 |
| Final Presentations | 5/6 |

Figure 4: The timeline for the overall project. Green-labeled milestones are part of the "Minimum" deliverables. Blue-labeled milestones are part of the "Expected" deliverables. Red-labeled milestones are part of the "Maximum" deliverables. Grey-labeled milestones are class reports and presentations.

In general, prior to code being written, either a test set or dataset will be constructed or obtained. After code is written, the code will be tested using the test set or dataset and the results will be recorded. By March 15, all deliverables of the identification detection and extraction should be completed. By the end of March, all deliverables of the data monitor detection and deformation should be completed. By the end of April, the algorithms should be fully incorporated into the smart glasses framework.

# 8 Roles and Responsibilities

## 8.1 Team Members

- **Robert Huang** (rhuang22@jhu.edu)
  *Undergraduate Senior, JHU Department of Biomedical Engineering*
  Sole responsibility for all tasks required in this project.

## 8.2 Team Mentors

- **Dr. Nick Dalesio** (nick.dalesio@mdairsupport.com)
  *JHH Anesthesiologist, MD Airsupport Co-Founder*
  Main Company Contact/Request, Dataset Generation, Network Architecture
- **Dr. Laeben Lester** (laeben.lester@mdairsupport.com)
  *JHH Anesthesiologist, Emergency Physician, and MD Airsupport Co-Founder*
  Headset Sample, EMS Experience
- **Dr. Mathias Unberath** (unberath@jhu.edu)
  *Professor, JHU Department of Computer Science*
  Advice for Algorithm Optimization

# 9 Management Plans

## 9.1 Meetings

There will be two weekly meetings to update mentors on the project's progress and to receive feedback and advice.

- General meeting to update and receive feedback from MD AirSupport on Friday 9AM, weekly.
- Imaging meeting with Dr. Mathias Unberath to update and ask questions about project on Friday 2PM, weekly.

## 9.2 Platforms

Many programs will be used throughout the project.

- **Communication:** Zoom will be used to have live meetings remotely. Email will be used to schedule the meetings and to discuss small topics.
- **Code Documentation:** Code and documentation will be stored on GitHub
- **Writing Reports:** Reports will be written using LaTeXon OverLeaf.
- **Document Storage:** All deliverables and presentations will be stored on the CIIS Wiki page made for this project.

## 10 Reading List

Listed below are references related to the various frameworks and algorithms that will be implemented. Above each reference, in bold, is a brief description of each article.

- **Description and Construction of the YOLO Deep Learning Framework**
  Redmon, Joseph, et al. "You Only Look Once: Unified, Real-Time Object Detection." *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, doi:10.1109/cvpr.2016.91.
- **Description and Construction of the Fast YOLO Framework**
  Shaifee, Mohammad Javad, et al. "Fast YOLO: A Fast You Only Look Once System for Real-Time Embedded Object Detection in Video." *Journal of Computational Vision and Imaging Systems*, vol. 3, no. 1, 2017, doi:10.15353/vsnl.v3i1.171.
- **Description and Construction of the Deep Learning Framework for Character Recognition**
  Wojna, Zbigniew, et al. "Attention-Based Extraction of Structured Information from Street View Imagery." *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 2017, doi:10.1109/icdar.2017.143.
- **Existing Program for Identity Card Reading**
  Llados, J., et al. "ICAR: Identity Card Automatic Reader." *Proceedings of Sixth International Conference on Document Analysis and Recognition*, doi:10.1109/icdar.2001.953834.
- **Blob Detection and Deformation Computer Vision Algorithm**
  Mikolajczyk K, Schmid C. Scale affine invariant interest point detectors. *International Journal on Computer Vision.* 2004;60:63. doi: 10.1023/B:VISI.0000027790.02288.f2.

## References


[1] Vuzix Corporation. Vuzix smart glasses at the chi mei medical center, taiwan. Technical report, 2020.

[2] Laeben Lester. Inquiry into ems documentation times, 2021.

[3] GreenLight Medical. Standardizing medical devices: Value analysis, 2020.

[4] Henning Muller Antoine Widmer Roger Schaer, Thomaz Melly. Using smart glasses in medical emergency situations, a qualitative pilot study. *2016 IEEE Wireless Health (WH)*, 2016.

[5] Radosveta Wells Stormy Monks Scott Crawford, Igor Kushner. Electronic health record documentation times among emergency medicine trainees. *Perspect Health Inf Manag*, 16, 2019.