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

Computer Integrated Surgery II
Spring, 2021
Robert Huang, under the auspices of Dr. Nick Dalesio, Dr. Laeben Lester, and Dr. Mathias Unberath

Introduction

- **Implemented detection and optical character recognition (OCR) algorithms** for the extraction of information from driver licenses and the Zoll and LIFEPAK vital monitors using smart glasses video feed.
- **Increased accuracy of extracted data** by using weighted frame averaging algorithm.

This was done to improve on-the-field outcomes by **moving time spent on documentation to patient care**, and by **increasing remote physician confidence and treatment speed**.

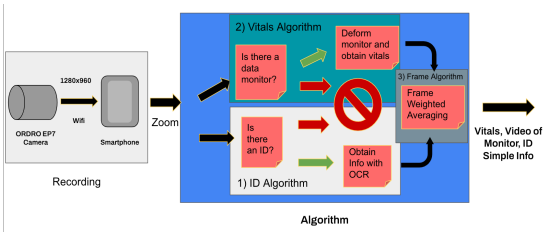
The Problem

- Smart glasses have only recently been introduced within healthcare settings to record and stream medical procedures, [1] and **there is room for AI and Computer Vision**.
- In on-the-field emergency medicine, **1-10 minutes can be spent obtaining simple information** (name, birth date, address, ID#). [2]
- Without recording vitals**, 43.4% of information can be lost**. [3] Furthermore, **physicians are 2-3x more confident and give treatments 2-3x faster** when they can confirm vitals for themselves. [3]

The Solution

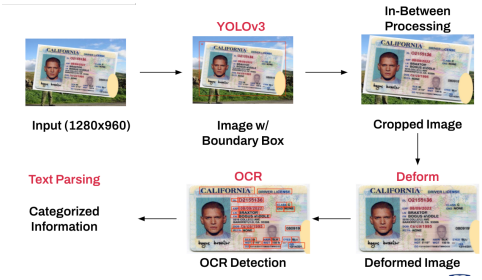
Detection and Extraction Algorithm

- Zoom recording is fed into three components in the algorithm: the ID extraction, the vitals extraction, and the frame averaging algorithm



Individual Detection and Extraction Algorithms

- Used YOLOv3 object detection, Hough Transform Edge detection and deformation, and Tesseract OCR (w/ some localization) to extract data from IDs and vitals.



Weighted Frame Averaging Algorithm

- Implemented weight frame averaging with inspiration from work by Petrova. [4]
- Frames that are in-focus contribute more to a running estimation of data.

Results

Component Tests

- Representative images were fed in, and accuracy was determined as: detected/total, true transform–estimated transform, true words/total words.
- 38ms per frame is too slow for 60 fps camera.

Step	Target Accuracy (or Loss)	Recorded Accuracy (or Loss)	Target Speed	Recorded Speed
YOLOv3	96%	99%	8ms	30ms
Deformation	<0.01	0.008	2ms	2ms
OCR and Location Parsing	96%	93%	3ms	6ms
Overall	88.8%	89.28%	16ms	38ms

Frame Averaging Tests

- Fed representative videos into full algorithm. Frame averaging increases accuracies.

Video	Target Accuracy (or Loss)	Recorded Accuracy (or Loss)
Driver Licenses	95%	94% -> 95%
Vitals Information	95%	~89% -> 99%

Future Work

- Incorporate more types of IDs into algorithm
- Incorporate more types of vitals monitors
- Implement the algorithm into the cloud-based workflow using live smart glasses feed.

Lessons Learned

- Use object detection, deformation, character recognition, and frame averaging algorithms.
- Data augment appropriately for the task.
- Construct tests to evaluate algorithm performance.

Acknowledgements

- We would like to thank Dr. Russell Taylor for his weekly feedback on this project, and Adam Toll and Mark Fedor for providing resources and feedback on behalf of MDAirSupport

References

[1] Vuzix Corporation. Vuzix Corporation, 2020, *VUZIX SMART GLASSES AT THE CHI MEI MEDICAL CENTER, TAIWAN*, ss-usa.s3.amazonaws.com/c/308483104/media/21105f5a523ce21ce43889049199725/Vuzix-Chi-Mei-Medical-Case-Study-2020.pdf.

[2] Crawford S, Kushner I, Wells R, Monks S. "Electronic health record documentation times among emergency medicine trainees." *Perspect Health Inf Manag*, 2019;16:1f.

[3] Schaer, et al. "Using Smart Glasses in Medical Emergency Situations, a Qualitative Pilot Study." *2016 IEEE Wireless Health (WH)*, 2016, doi:10.1109/wh.2016.7764556.

[4] O. Petrova, K. Bulatov, V. Arlazarov, V. Arlazarov, Weighted combination of per-frame recognition results for text recognition in a video stream, Computer Optics. 45 (2021) 77–89. doi:10.18287/2412-6179-co-795.

Datasets

License Dataset

Dataset Characteristics:

- **23834 Data Augmented** Images of ID
  - USA IDs transformed and brightness-adjusted and pasted onto empty backgrounds.
  - Add geometry (pentagon, circle, or oval) of 'skin' and white/grey colors to emulate obstructions.
- **11723 Natural Images** of ID
- **7209 Negative Images**
  - Images of Potential Confusers: Smartphones, Billboards, Books, etc.

[7] [8]

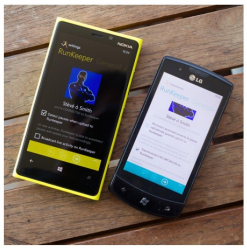
ID transformed and 'pasted' on background.

[7] Gould, Stephen, Richard Fulton, and Daphne Koller. 2009. "Decomposing a Scene into Geometric and Semantically Consistent Regions." *2009 IEEE 12th International Conference on Computer Vision*. doi:10.1109/iccv.2009.5459211.
[8] Shaw, Gabbi, and Frank Olito. 2020. "What a Driver's License Looks like in Every State." *Insider*. Insider. January 21. https://www.insider.com/what-drivers-license-looks-like-in-every-state.

License Dataset

MIDV-500 Natural Images of Non-US IDs [9]

Negative Images of potential confusers [10]



[9] V. V. Arlazarov, K. Bulatov, T. Chernov and V. L. Arlazarov, "A dataset for identity documents analysis and recognition on mobile devices in video stream", *Comput. Opt.*, vol. 43, no. 5, pp. 818-824, 2019.
[10] Kuznetsova, Alina, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, et al. 2020. "The Open Images Dataset V4." *International Journal of Computer Vision* 128 (7): 1956–81. doi:10.1007/s11263-020-01316-z.

Vitals Monitor Dataset

Monitors:

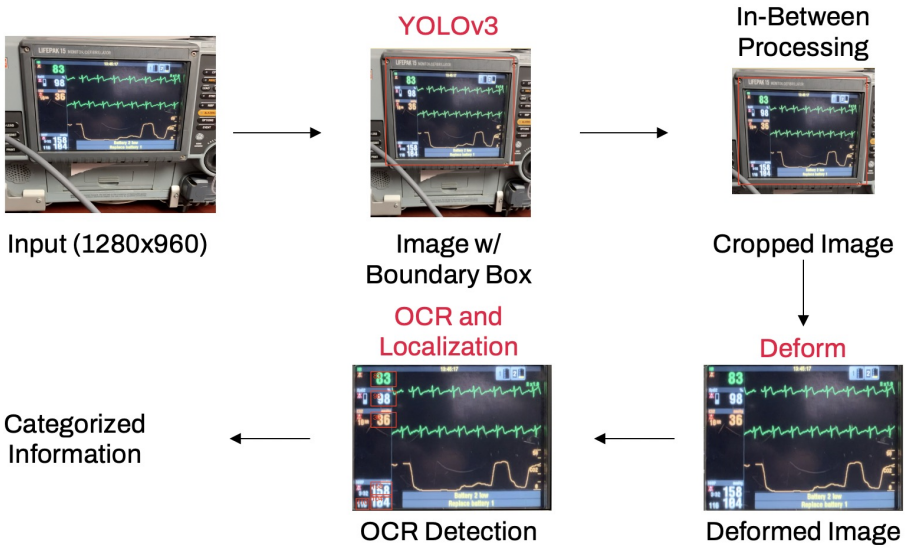
- 3240 Positive Zoll Images
  - Homography, Color, and Brightness Augmented
- 5620 LIFEPAK Images
  - Homography, Color, and Brightness Augmented
- 7209 Negative Images
  - Potential Confusers: Smartphones, Billboards, Books, etc.

Other Figures

Performance - Identification

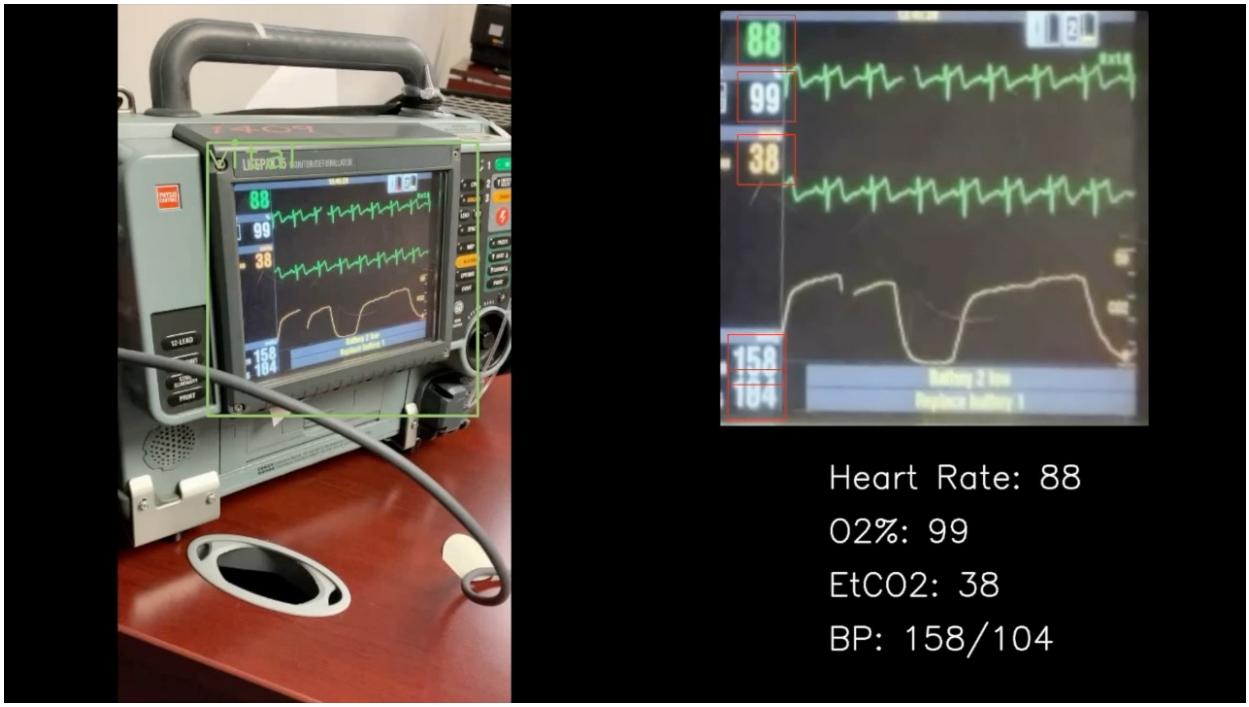
Step	Target Accuracy (or Loss)	Recorded Accuracy (or Loss)	Target Speed	Recorded Speed
YOLOv3	95%	99%	8ms	30ms
Deformation	<0.01	0.009	2ms	1ms
OCR	95%	96%	3ms	6ms
Text Parsing	95%	99%	3ms	<1ms
Overall	85.7%	94%	16ms	37ms

Technical Approach - Vitals Monitor



Performance - Vitals Monitor

Step	Target Accuracy (or Loss)	Recorded Accuracy (or Loss)	Target Speed	Recorded Speed
YOLOv3	96%	99%	8ms	30ms
Deformation	<0.01	0.008	2ms	2ms
OCR and Location Parsing	96%	93%	3ms	6ms
Overall	88.8%	89.28%	16ms	38ms


88
99
38
158
104
Heart Rate: 88
02%: 99
EtCO2: 38
BP: 158/104