AI-fueled FPV Goggles with Mobile Device and VR Headset

Presenter

Sydney Brown
Barnard College
skb2183@barnard.edu

Mentor

Qing Su
Georgia State University
qsu3@student.gsu.edu

Advisor

Dr. Shihao Ji
Georgia State University
sji3@gsu.edu

Context

- Drone technology is used for professional and recreational purposes
- FPV Goggles allow users to immerse themselves in a drone's view
- Existing FPV goggles are expensive additional hardware
- Creating robust FPV goggles on mobile would allow casual users to immersively view drone video

Research Problem

How to access and convert monocular video footage from a drone into stereo video and display it on a mobile phone device?

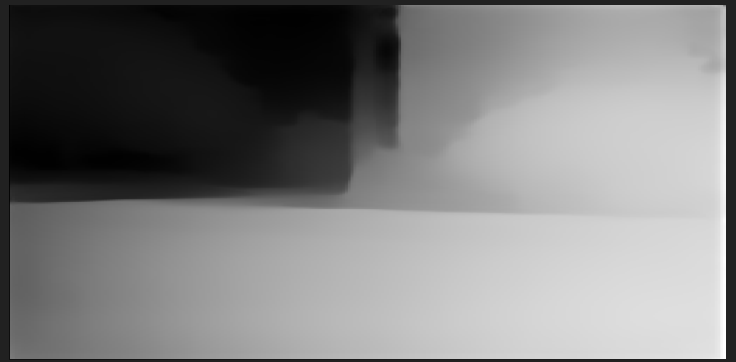
Turning Monocular Video into Stereo Video

- Principle of parallax in human vision
- VR uses parallax to create perception of depth
- One way to create stereo video involves filming at multiple views
- With only monocular video, must extrapolate depth information
- Can use depth information to construct stereo video

Monocular Depth Estimation

- Topic of research in area of Computer Vision
- Ill-posed problem of generating a depth map of a single RGB image [1]
- Solutions often computationally expensive, involving large CNNs [1]

Image frame

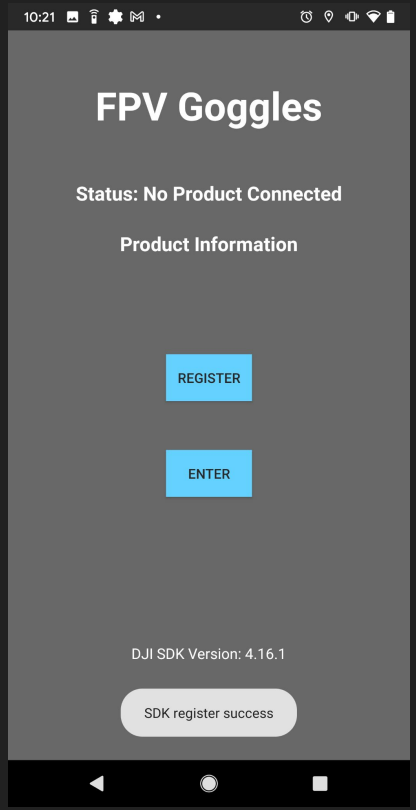


Depth map

Implementation

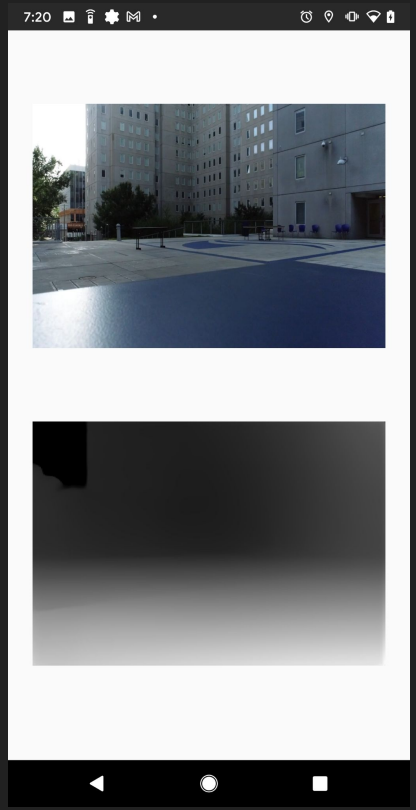
Set Up and Integration with DJI SDK

- Hardware used for testing:
  - DJI Phantom 4 Pro V2
  - Android Pixel 2 XL
  - Google Daydream
- Create app registering DJI SDK and connecting to drone [2]
- Access and display the live video feed from the drone [2]



Implementing Depth Estimation

- Access video frames from the drone’s video feed
- Use TensorFlow Lite to load and run MiDaS depth estimation model [3]
- Process input to required format for Tensors [4]
- Process output into a depth map represented by a grayscale image



Creating Second View

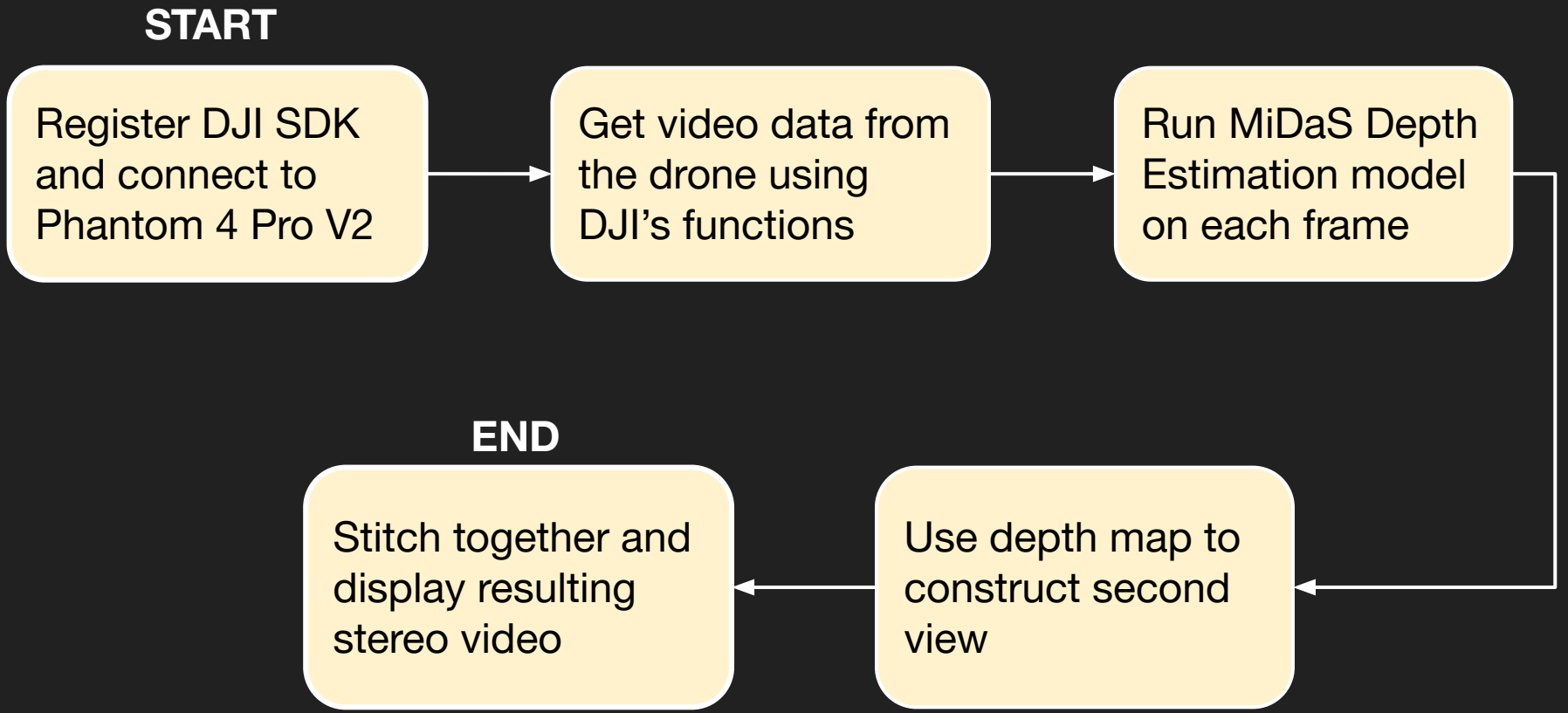
- Attempted to implement methods for creating the stereo pair on mobile
- First method split image into layers based on depth and shifted them [5]
- Second shifted individual pixels according to depth [6]
- Issues with resulting frames
- Not feasible to carry out this processing for 30 fps

Original frame

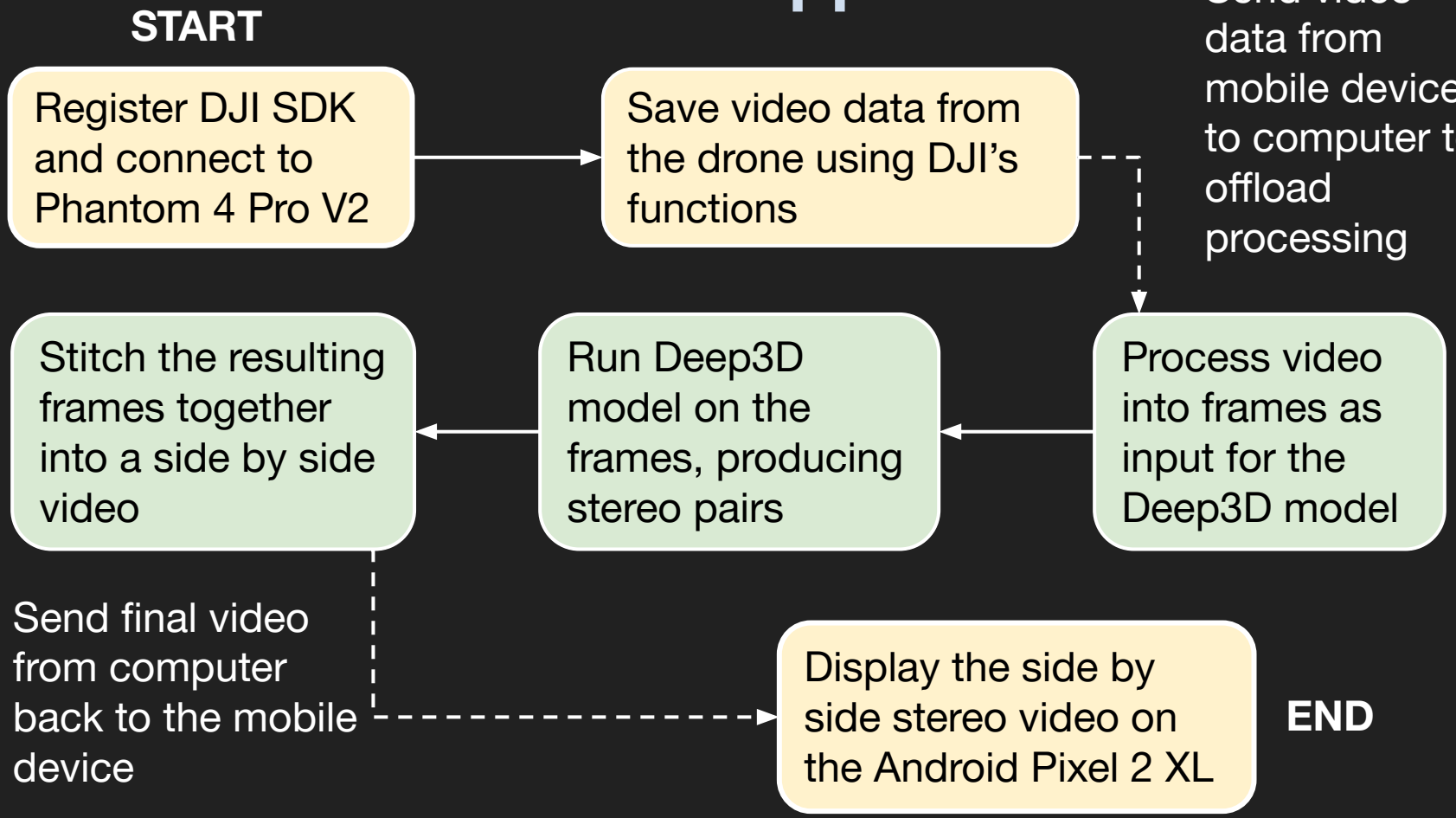
Other view using first method

Other view using second method

Initial Approach



Revised Approach



Offloading Video and Processing

- Download video data from drone
- Deep3D – model for producing stereo pair from single image [7]
- Process video into frames on the computer [8]
- Run Deep3D on the input frames [8]
- Stitch back together output as side by side stereo video [8]

Final Result

Conclusions and Future Work

- Implemented FPV Goggles on mobile device
- Due to expense of image processing and model inference, decided to offload this work
- Future work could develop lightweight, efficient model for mobile
- Potential to implement Deep3D as a TensorFlow Lite model to run directly on mobile
- Improve rendering of final result on mobile

References

1. C. Q. Zhao, Q. Y. Sun, C. Z. Zhang, Y. Tang, and F. Qian, “Monocular depth estimation based on deep learning: An overview,” Science China Technological Sciences, vol. 63, 2020.

2. oliverou, H. Vo, mrjzhao, and DJI-William, Android FPV Demo. DJI, 2022. [Online]. Available: https://github.com/DJI-Mobile-SDKTutorials/Android-FPVDemo.

3. R. Ranftl, K. Lasinger, D. Hafner, K. Schindler, and V. Koltun, “Towards robust monocular depth estimation: Mixing datasets for zero-shot crossdataset transfer,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 3, pp. 1623–1637, 2022.

4. A. Shekhar, “Android tensorflow lite example,” 2019. [Online]. Available: https://github.com/amitshekhariitbhu/Android-TensorFlowLite-Example.

5. strikeraryu, “Parallax image,” Medium, 2020. [Online]. Available: https://medium.com/analytics-vidhya/parallax-images-14e92ebb1bae.

6. J. Zimandl, “Stereo photo from depth photo,” STRV, 2019. [Online]. Available: https://www.strv.com/blog/stereo-photo-from-depthphoto-engineering-ios.

7. J. Xie, R. Girshick, and A. Farhadi, “Deep3d: Fully automatic 2dto-3d video conversion with deep convolutional neural networks,” in Computer Vision – ECCV 2016, vol. 9908 LNCS, pp. 842–857, Springer International Publishing, 2016.

8. HypoX64, “Deep3d,” 2022. [Online]. Available: https://github.com/HypoX64/Deep3D.

Thank You

NSF Grant #1852516

Presenter

Sydney Brown
Barnard College
skb2183@barnard.edu

Mentor

Qing Su
Georgia State University
qsu3@student.gsu.edu

Advisor

Dr. Shihao Ji
Georgia State University
sji3@gsu.edu

AI-fueled FPV Goggles with Mobile Device and VR Headset

Sydney Brown
Barnard College
skb2183@barnard.edu

Qing Su
Georgia State University
qsu3@student.gsu.edu

Shihao Ji
Georgia State University
sji@gsu.edu

NSF Grant #1852516

Introduction

- FPV Goggles provide an immersive format of viewing drone footage
- True stereo video includes two views which account for parallax
- Neural networks can be used to produce depth maps from a single RGB image
- How can we create FPV goggles on a mobile device, accounting for parallax?

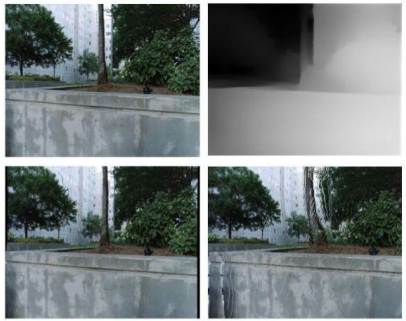
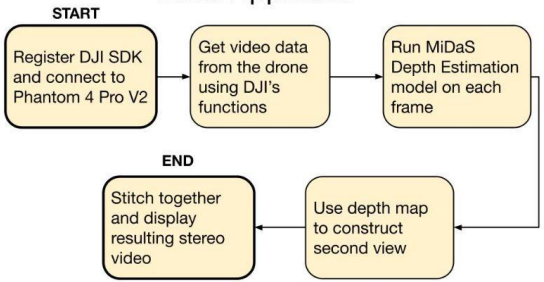


Figure 1. First row left to right: original image from drone, generated depth map Second row: two versions of creating second view using different shifting methods

Implementation

Initial Approach



Revised Approach

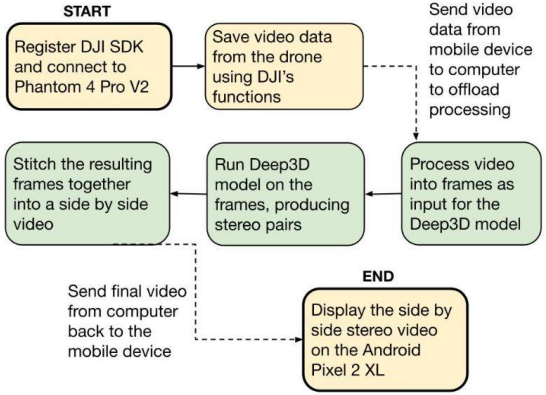


Figure 2. Some resulting stereo frame pairs generated from Deep3D

Conclusion

- Due to the processing power of the mobile phone, approach shifted to offloading model inference and video processing
- The final result (example frames displayed in Figure 2) can be displayed on mobile and viewed on the VR headset
- Future work could create lightweight, efficient depth estimation models for mobile
- Could implement Deep3D in TensorFlow Lite and directly deploy the model on mobile