ACCELERATING OPTICAL FLOW AND STEREO DISPARITY ESTIMATION USING THE NVIDIA A100 OFA ENGINE

Eric Viscito

AGENDA

Optical Flow and Stereo Disparity Definition and Applications

What is Optical Flow? What is Stereo Disparity?

How are they used in computer vision applications?

OFA Engine – Motivation and Principles of Operation

Why did we build OFA? What can it do? How does it work?

Quality and Performance Metrics

How do we measure quality, and how does OFA do?

What throughput can OFA achieve?

Programming Flexibility

How can OFA be tailored to specific applications?

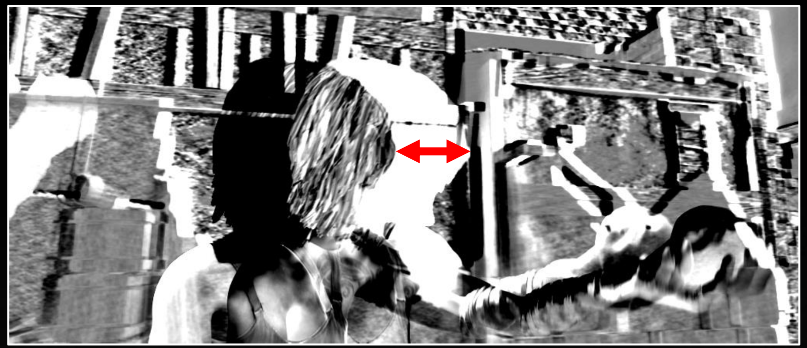
OPTICAL FLOW AND *STEREO DISPARITY* DEFINITION AND APPLICATIONS

STEREO DISPARITY

Depth from two parallel calibrated cameras

Left view

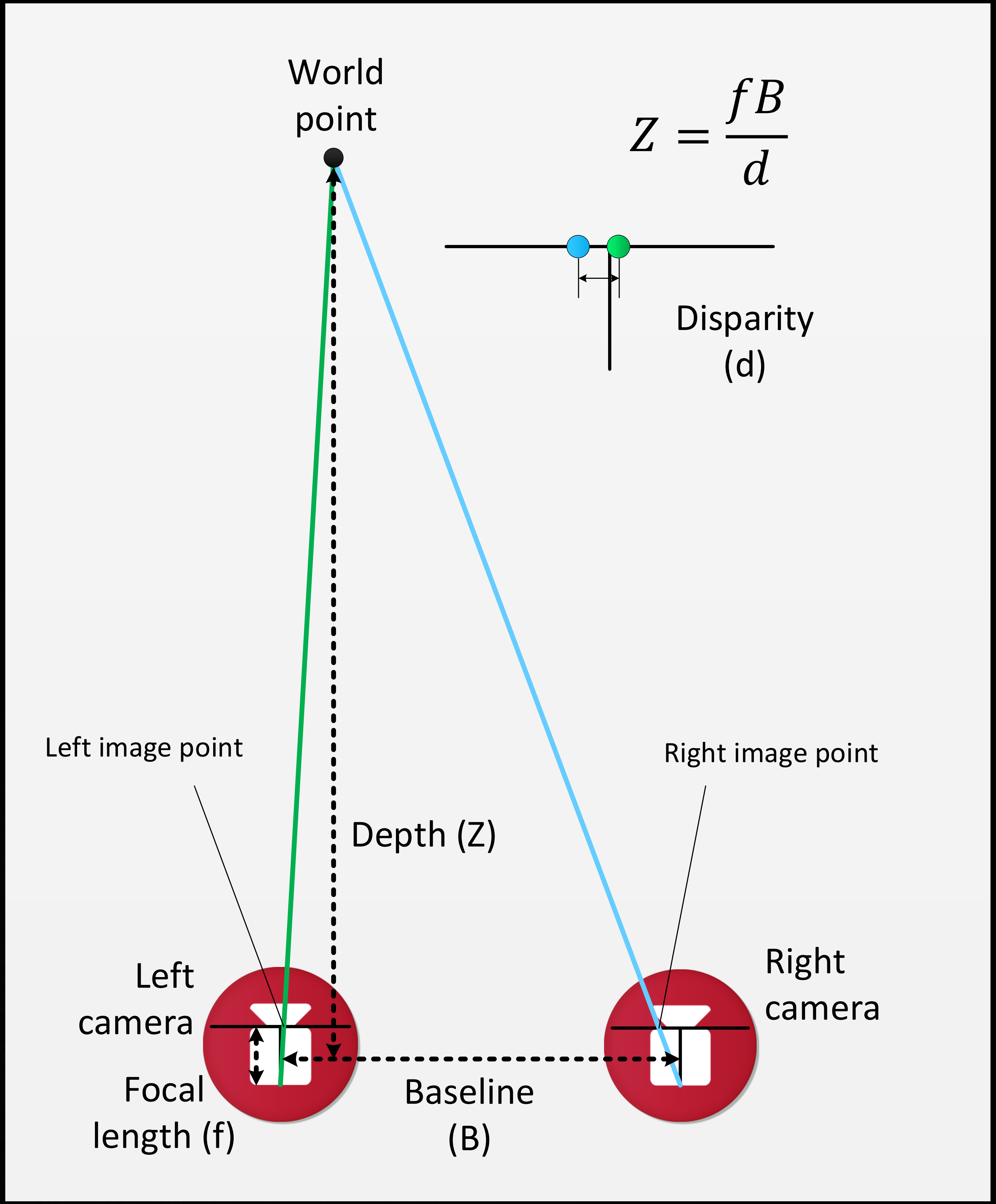
Right view



Disparity at one point

Dense disparity map

- Factors affecting stereo resolving power
  - Baseline and focal length
  - Camera field of view (30°, 60°, etc.) [lower is better]
  - Sensor resolution [larger better] and pixel size [smaller better]



OPTICAL FLOW

Movement on image plane between two views from a single camera

Animated GIF

VIRTUAL REALITY

Facebook 3D-360 video project

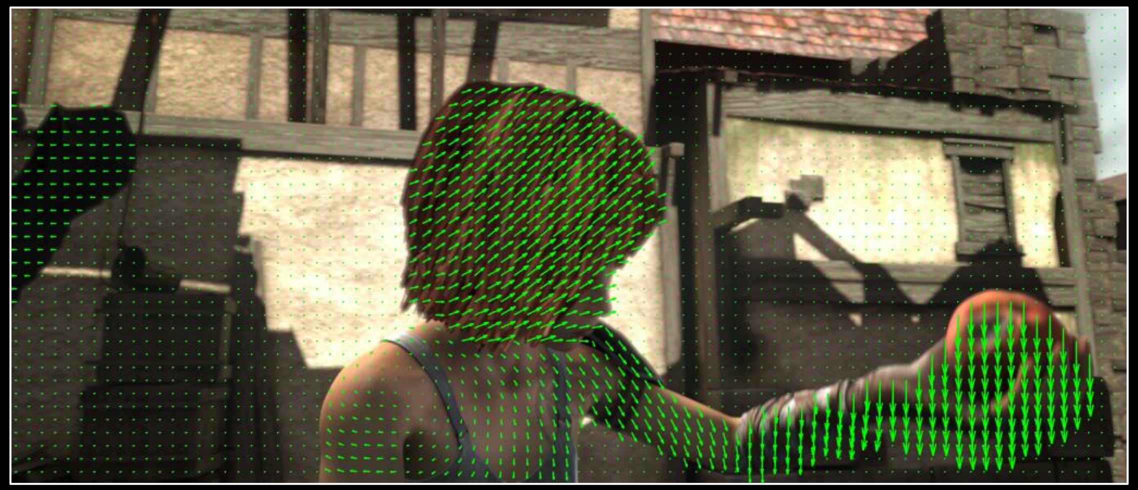
https://www.facebook.com/Engineering/videos/10154013275372200/

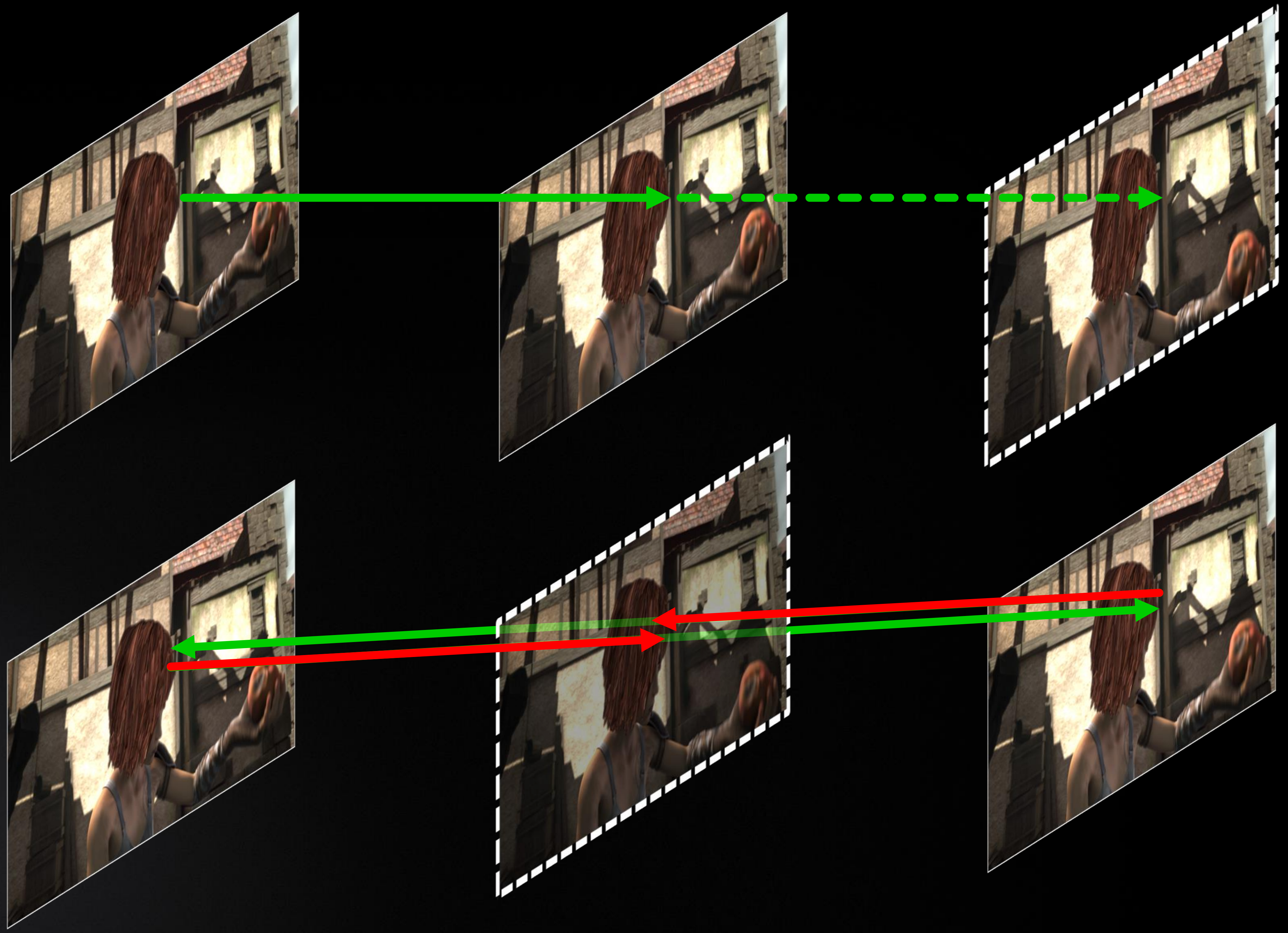
Code available: https://github.com/facebook/Surround360

IMPROVE GAMING EXPERIENCE

Frame rate upsampling

Interpolation vs Extrapolation





Oculus ASW uses flow-based extrapolation to maintain frame rate in hard-to-render scenes

Interpolation to increase frame rate is better when end-to-end delay is not as important

VIDEO UNDERSTANDING

OPTICAL FLOW IN VIDEO ACTION RECOGNITION

Two-stream CNN for Video Action Recognition

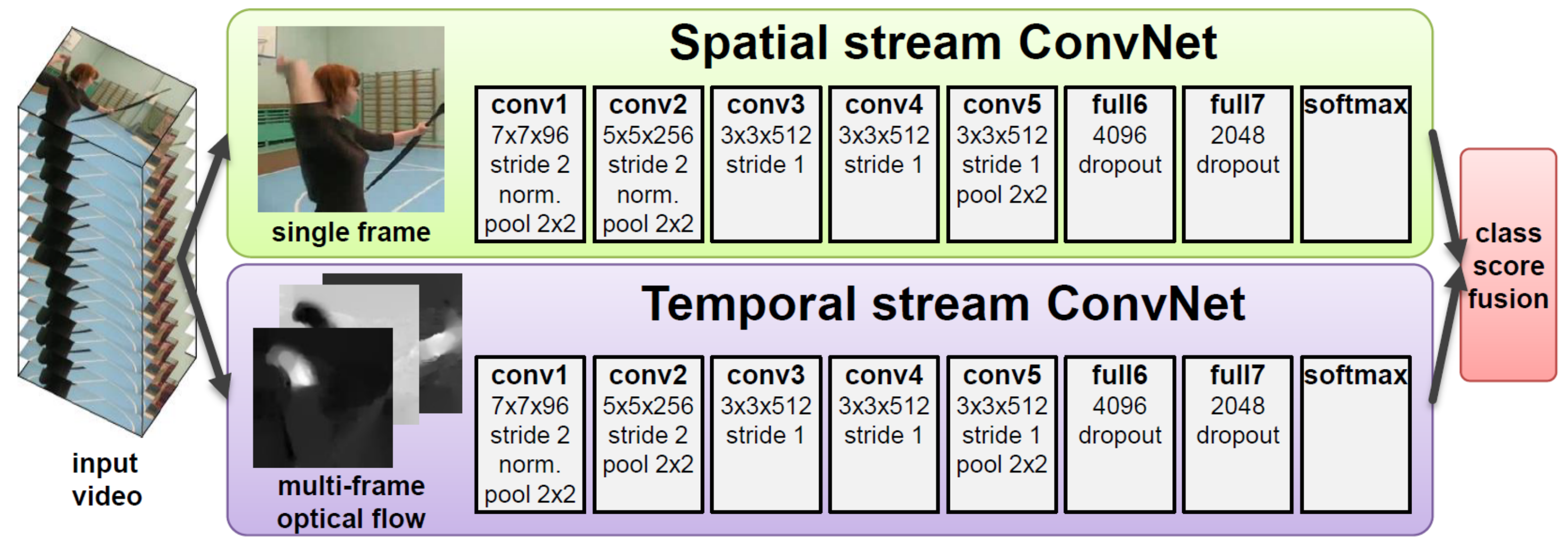


Figure 1: **Two-stream architecture for video classification.**

"Our experiments on two challenging datasets (UCF-101 and HMDB-51) show that the two recognition streams (image and optical flow) are complementary…"

Table 1: **Individual ConvNets accuracy on UCF-101 (split 1).**

(a) **Spatial ConvNet.**

Training setting	Dropout ratio	
	0.5	0.9
From scratch	42.5%	52.3%
Pre-trained + fine-tuning	70.8%	**72.8%**
Pre-trained + last layer	**72.7%**	59.9%

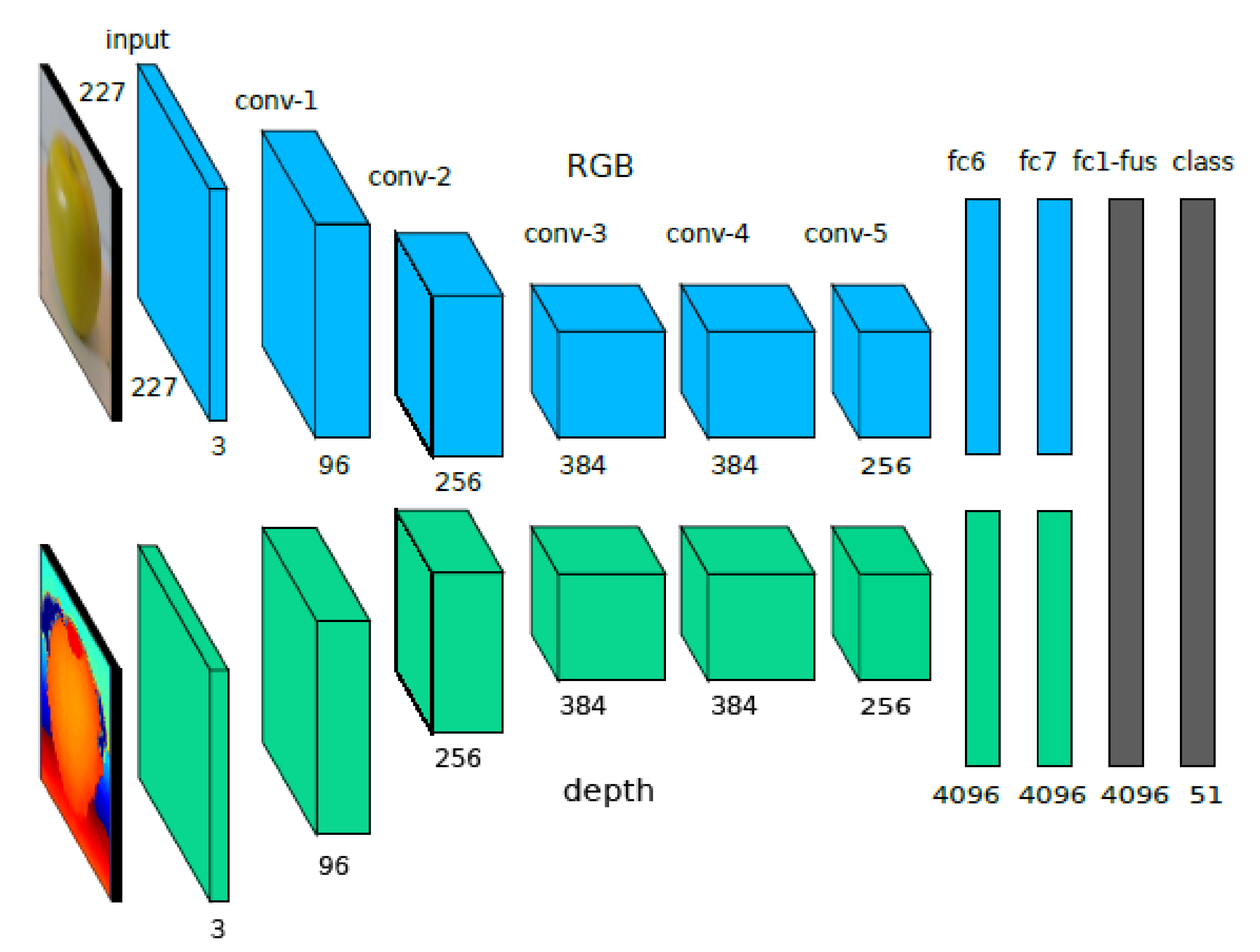
Table 3: **Two-stream ConvNet accuracy on UCF-101 (split 1).**

Spatial ConvNet	Temporal ConvNet	Fusion Method	Accuracy
Pre-trained + last layer	bi-directional	averaging	85.6%
Pre-trained + last layer	uni-directional	averaging	85.9%
Pre-trained + last layer	uni-directional, multi-task	averaging	86.2%
Pre-trained + last layer	uni-directional, multi-task	SVM	**87.0%**

From "*Two-Stream Convolutional Networks for Action Recognition in Videos*", Simonyan and Zisserman, 2014

STEREO DISPARITY IN OBJECT RECOGNITION

Two-stream CNN for RGB-D object recognition



Recognition accuracy improvement

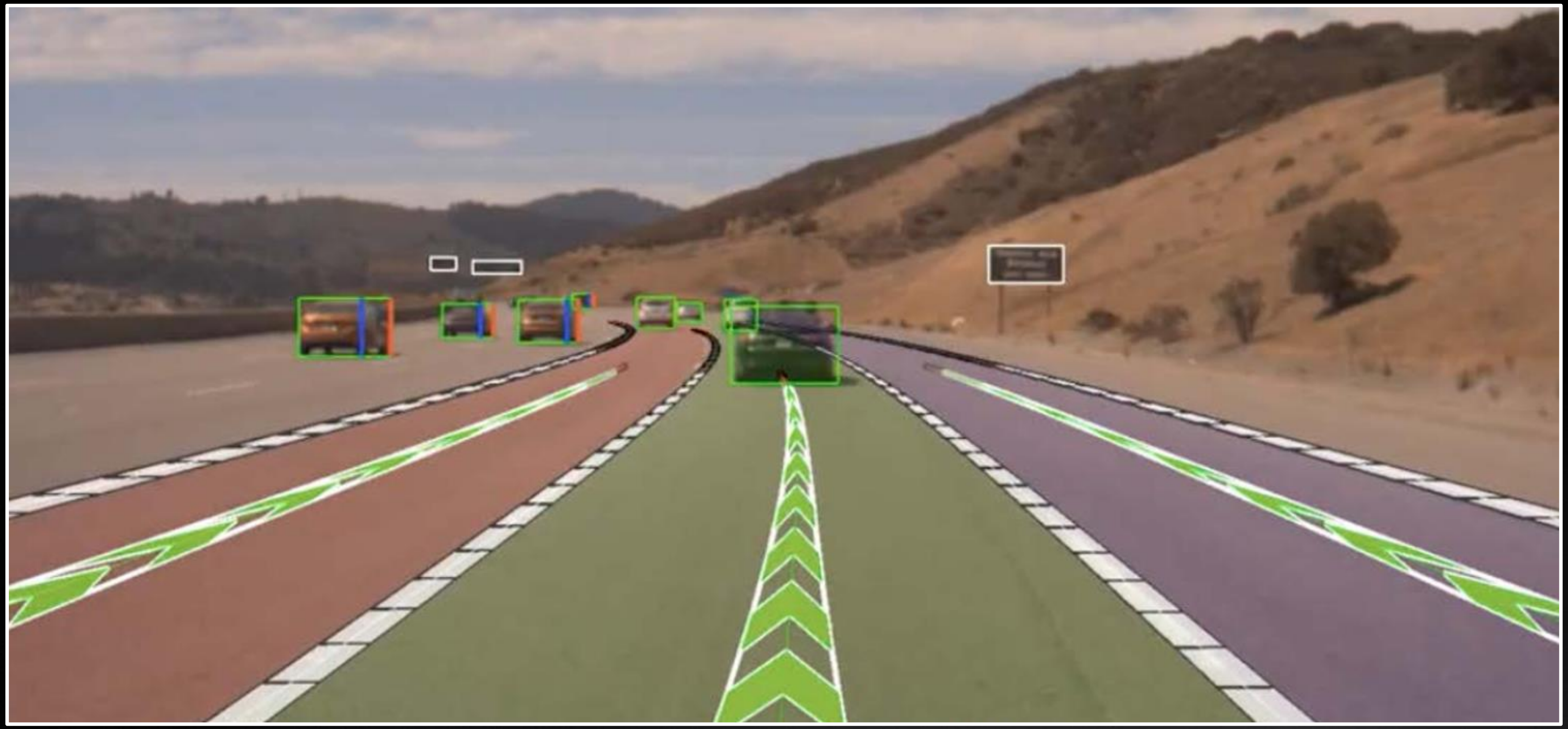
TABLE I: Comparisons of our fusion network with other approaches reported for the RGB-D dataset. Results are recognition accuracy in percent. Our multi-modal CNN outperforms all the previous approaches.

Method	RGB	Depth	RGB-D
Nonlinear SVM [15]	74.5 ± 3.1	64.7 ± 2.2	83.9 ± 3.5
HKDES [4]	76.1 ± 2.2	75.7 ± 2.6	84.1 ± 2.2
Kernel Desc. [14]	77.7 ± 1.9	78.8 ± 2.7	86.2 ± 2.1
CKM Desc. [3]	N/A	N/A	86.4 ± 2.3
CNN-RNN [22]	80.8 ± 4.2	78.9 ± 3.8	86.8 ± 3.3
Upgraded HMP [5]	82.4 ± 3.1	81.2 ± 2.3	87.5 ± 2.9
CaRFs [1]	N/A	N/A	88.1 ± 2.4
CNN Features [20]	83.1 ± 2.0	N/A	89.4 ± 1.3
Ours, Fus-CNN (HHA)	$\mathbf{84.1 \pm 2.7}$	$\mathbf{83.0 \pm 2.7}$	$\mathbf{91.0 \pm 1.9}$
Ours, Fus-CNN (jet)	$\mathbf{84.1 \pm 2.7}$	$\mathbf{83.8 \pm 2.7}$	$\mathbf{91.3 \pm 1.4}$

From "*Multimodal Deep Learning for Robust RGB-D Object Recognition*", Eitel *et al*, 2015

AUTOMOTIVE USAGE

Driver Assistance and Autonomous Driving



ADAS task examples

- Automatic emergency braking
- Lane keep assistance
- Hazard alert

Autonomous task examples

- Free space / driving lane detection
- Obstacle detection and identification
- Mapping and localization
- Path planning / driving policy

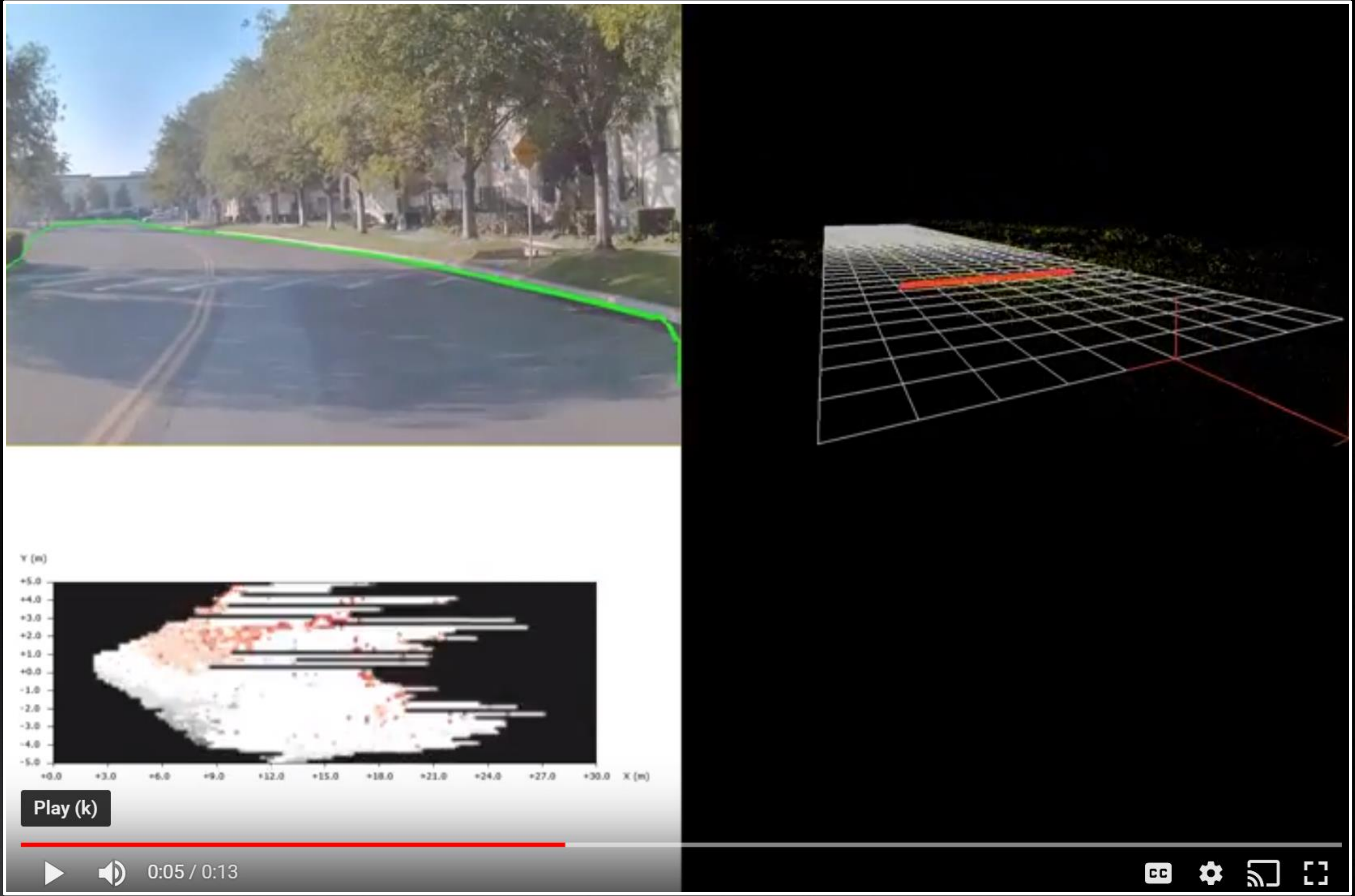
AUTOMOTIVE USAGE EXAMPLES

Road profile analysis

Potholes, bumps, hazardous objects, etc.

Hazardous object detection

Bump detection



ROBOTICS APPLICATIONS

Agricultural concept

Capture 3D scene with depth from stereo or structure-from-motion from mono camera

Analyze semantics using deep learning

Applications are many

- Inspection
- Targeted pest control
- Trimming
- Harvesting

Credit: StereoLabs (https://www.stereolabs.com/solutions/robotics/)

OFA ENGINE – MOTIVATION AND PRINCIPLES OF OPERATION

OFA = OPTICAL FLOW ACCELERATOR

Evolution of Nvidia Optical Flow / Stereo Hardware

ME-only
- Purely for compression
- Inside the NVENC encoder

SOFE
- Adapt ME-only for optical flow
- Turing, Xavier

OFA
- Dedicated HW for optical flow and stereo
- A100, Orin

MOTIVATION FOR OFA ENGINE

Quality: Significant improvement over SOFE

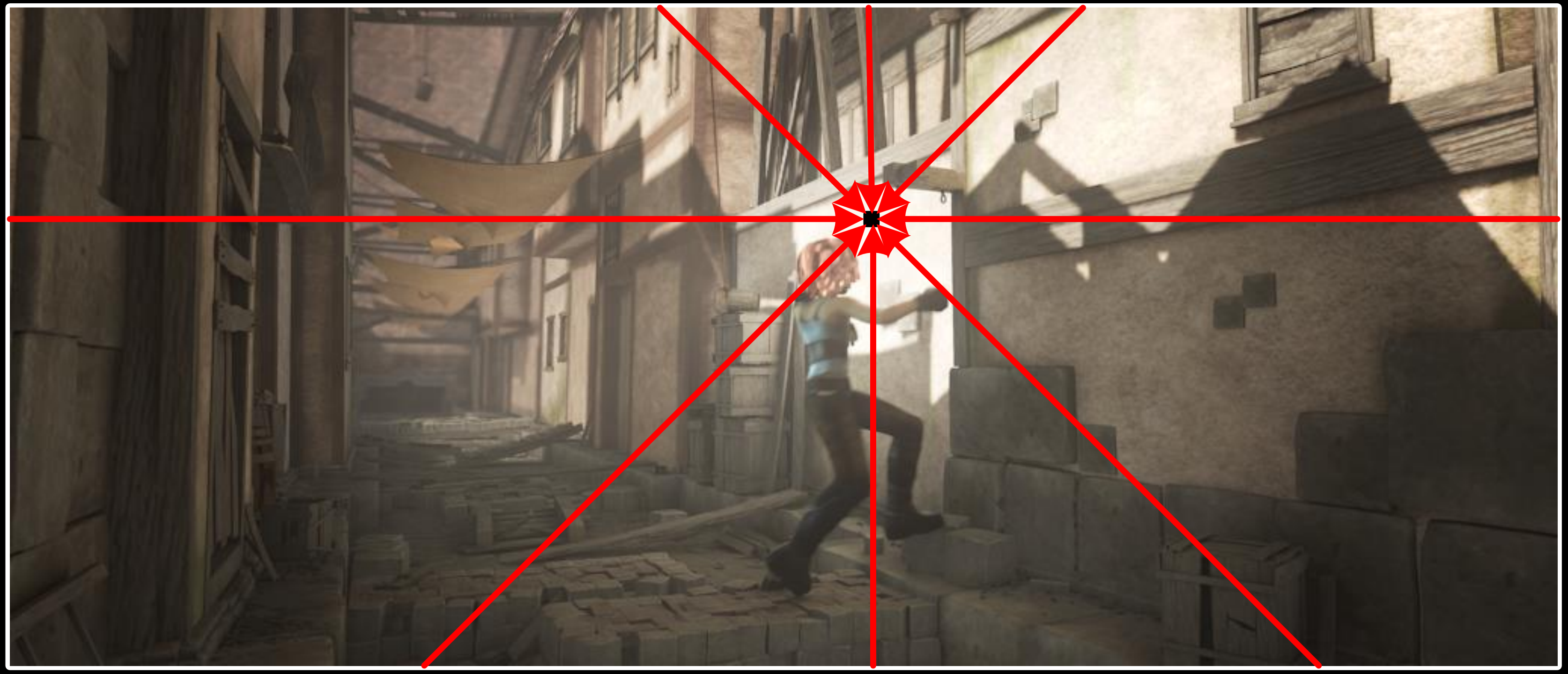
Performance: Match SOFE

Flexible application

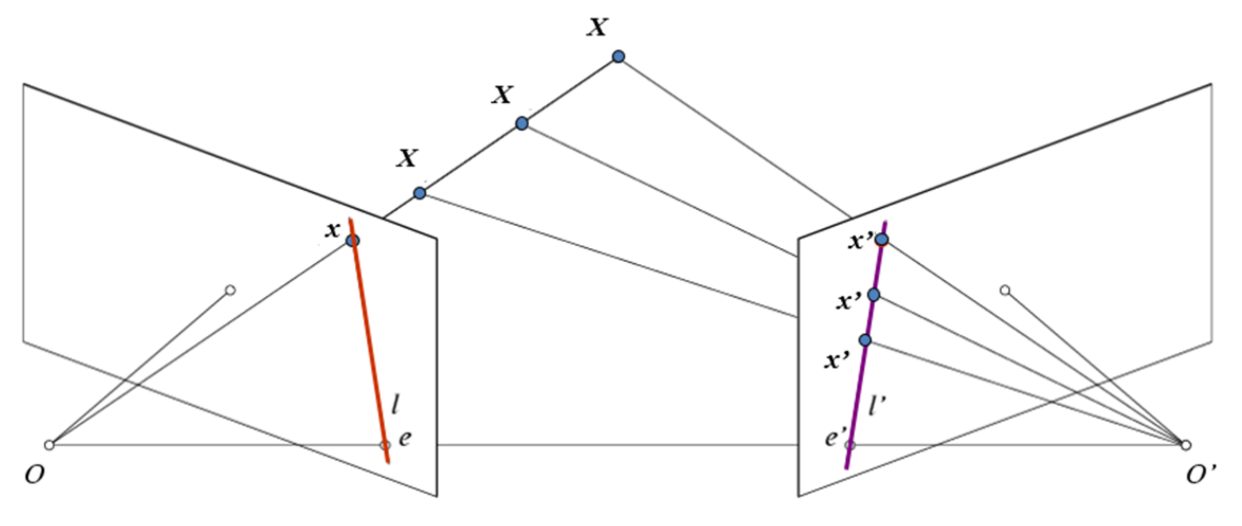
Support: Both Tegra and GPU

BASIS OF OFA ENGINE

Real-time motion / depth smoothing (SGM)

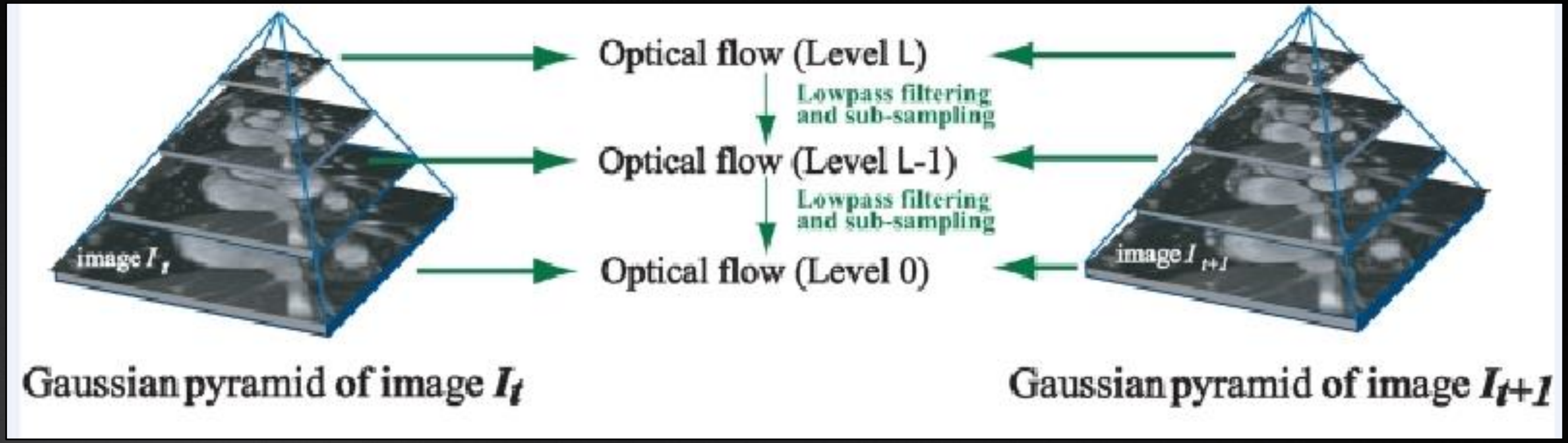


Rigid static world & inter-camera geometry (Epipolar)

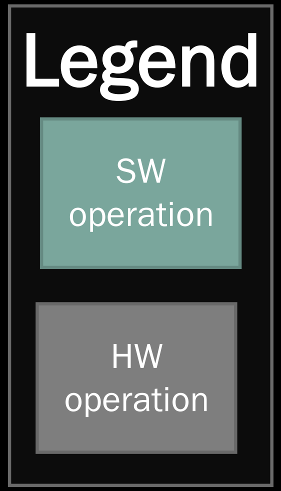


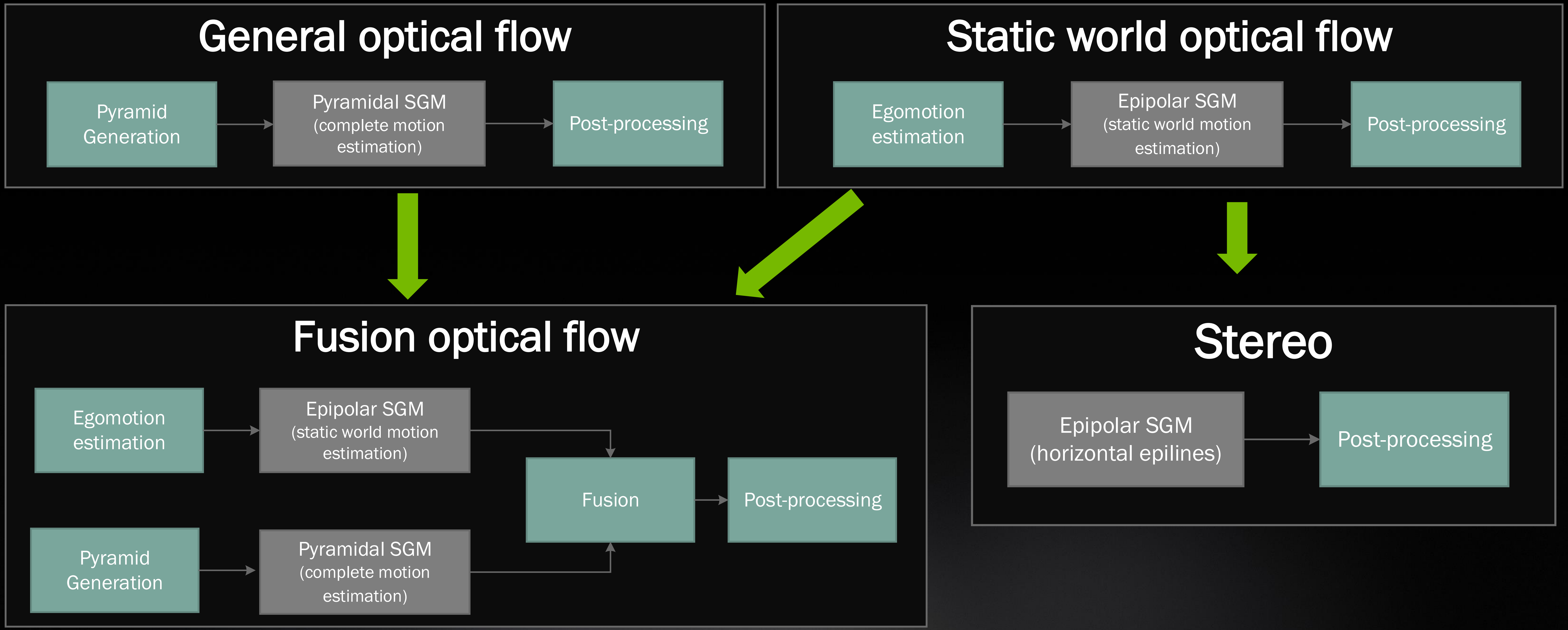
Optical flow w/o geometry and camera data

(Pyramidal SGM)



MODES OF OPERATION



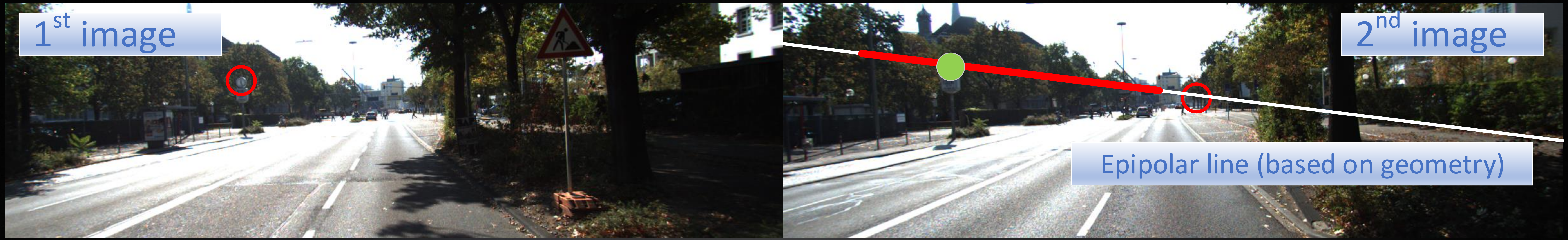


EPIPOLAR SGM FOR OPTICAL FLOW

Extending SGM to Optical Flow – Slide 1

Stereo case: matches appear along the same horizontal line

General two-camera case: matches appear along the epipolar line



EPIPOLAR GEOMETRY CALCULATION

Extending SGM to Optical Flow – Slide 1

I_0, I_1
Intrinsic matrix K →

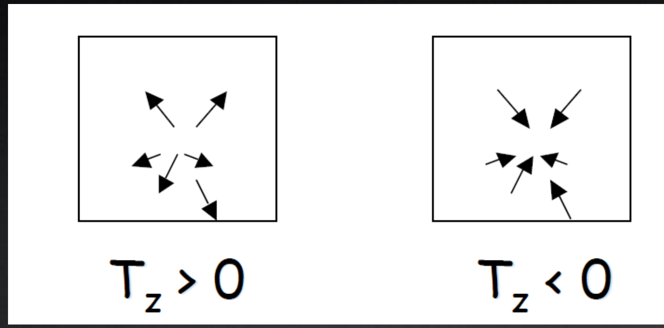
- Extract features
- ↓ Match features
- ↓ Estimate fundamental matrix
- ↓ Estimate rotation, calculate epipole
- ↓ Compute Homography
- ↓ Determine motion direction

← $F, H, \mathrm{e}', \mathrm{d}$

$$x'^T F x = 0$$

$$E = K' F K \ \rightarrow \mathrm{R}\ \&\ \mathrm{epipole}$$

$$\mathrm{H} = K R K'$$



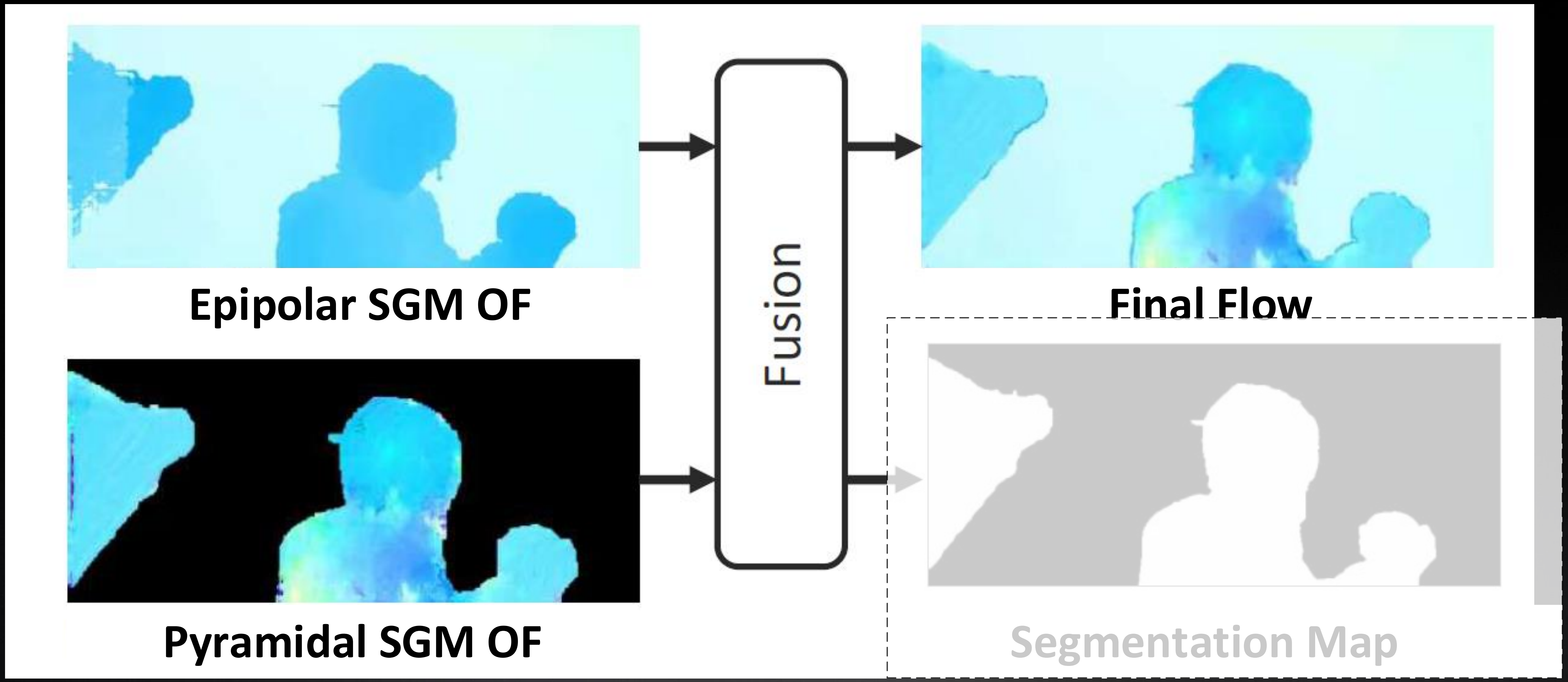
- Relies upon knowledge of camera intrinsic parameters
- Tricky to do robustly

Multiple View Geometry, Hartley and Zisserman

FUSION MODE

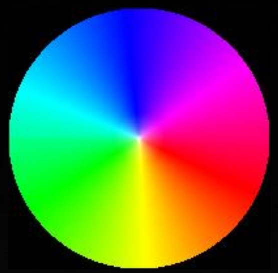
Step 1: Run OFA with epipolar & pyramidal SGM mode sequentially

Step 2: Fuse the two generated flow maps, using the corresponding cost maps or other data



QUALITY AND PERFORMANCE

OPTICAL FLOW VISUALIZATION

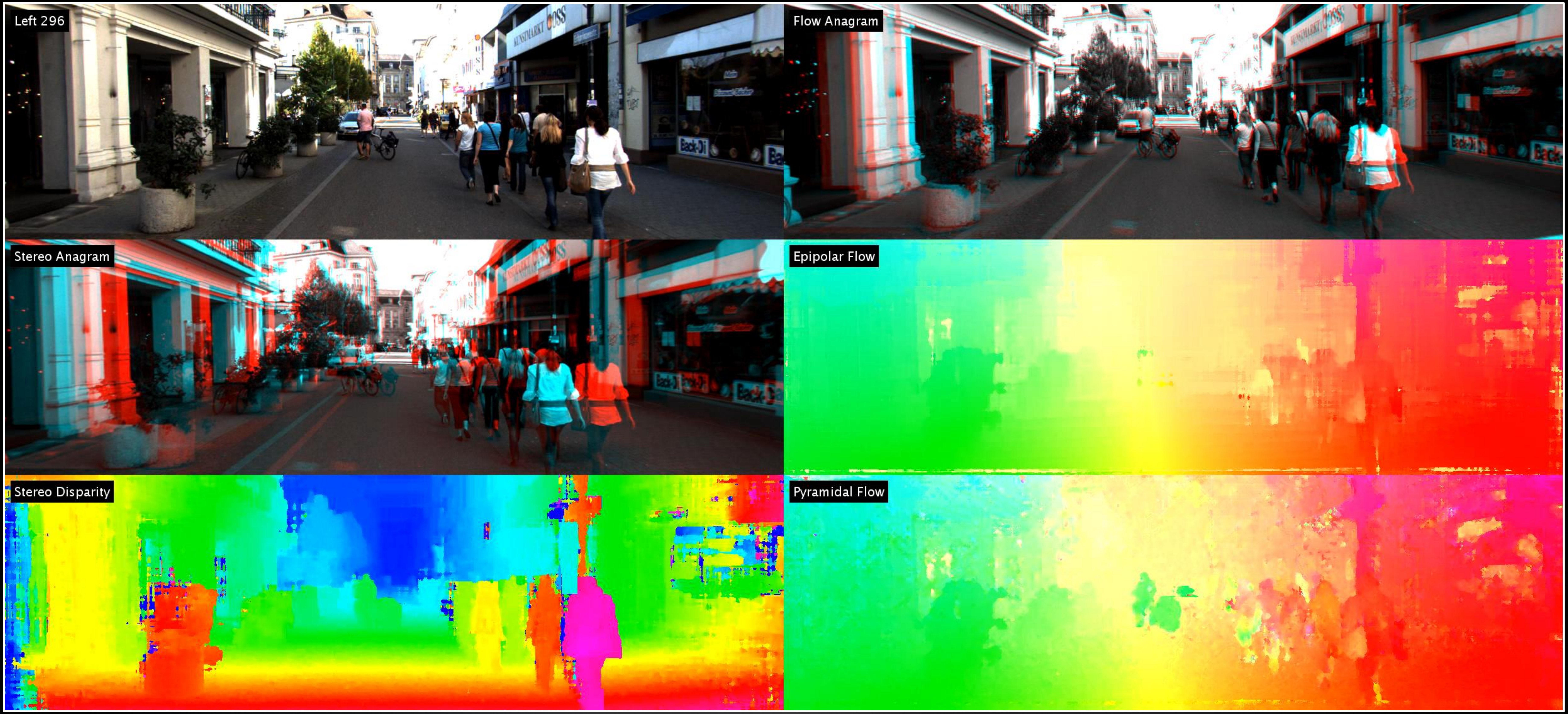


- Color Map
  - hue indicates direction of flow
  - saturation indicates the flow magnitude
  - Good for assessing smoothness and boundaries at full density

- Arrow map
  - Each arrow represents the flow vector at that point
  - Good for assessing accuracy, but at lower density

Example from MPI Sintel

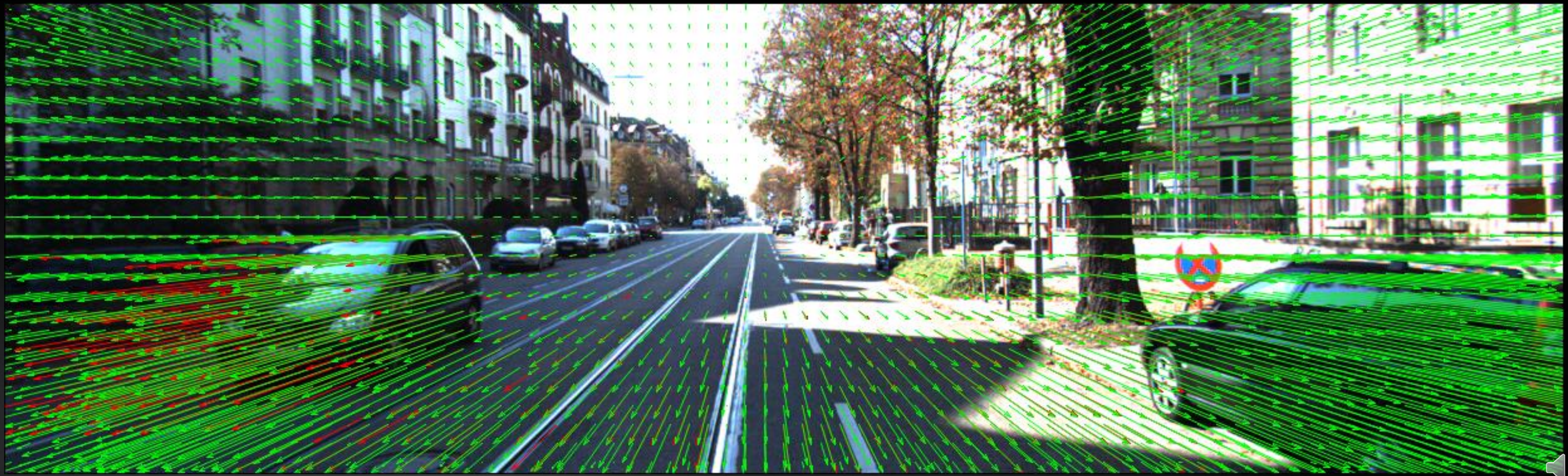
EXAMPLE FROM URBAN DRIVING SCENARIO (COLORIZED)



EXAMPLE FROM KITTI 2015 BENCHMARK (ARROWS)

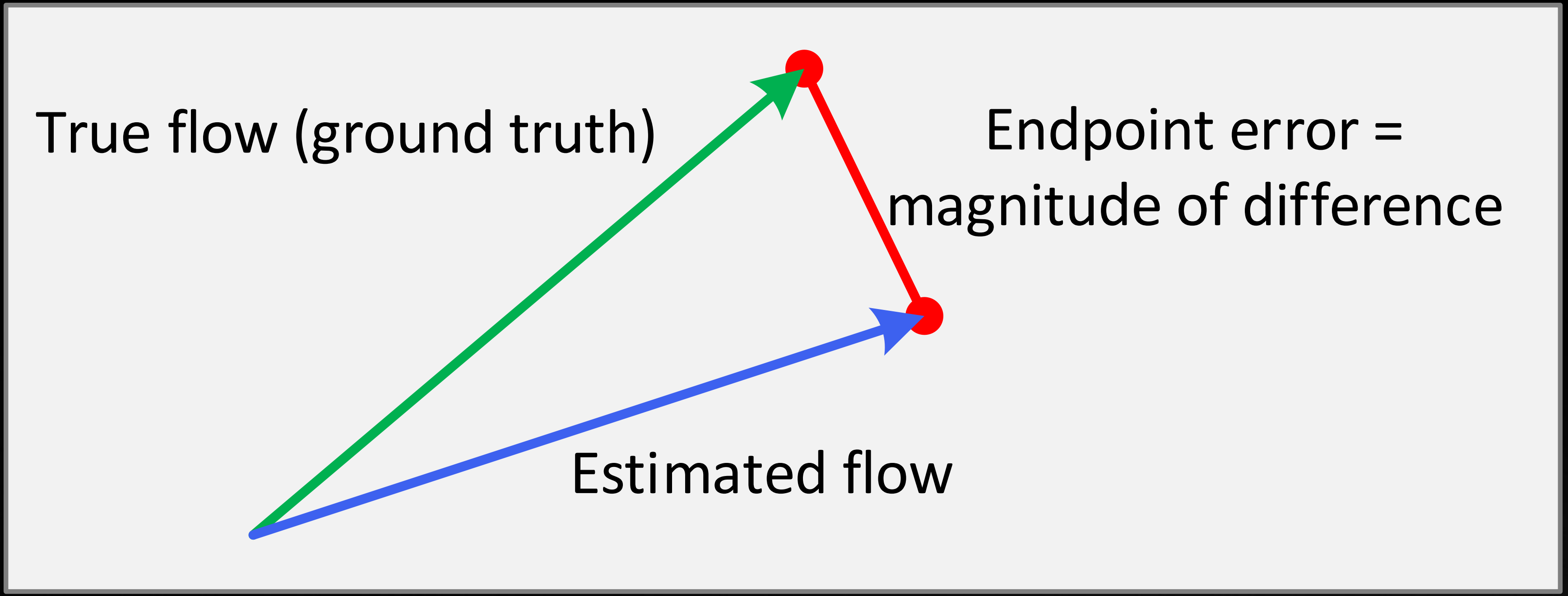
SOFE optical flow

OFA optical flow

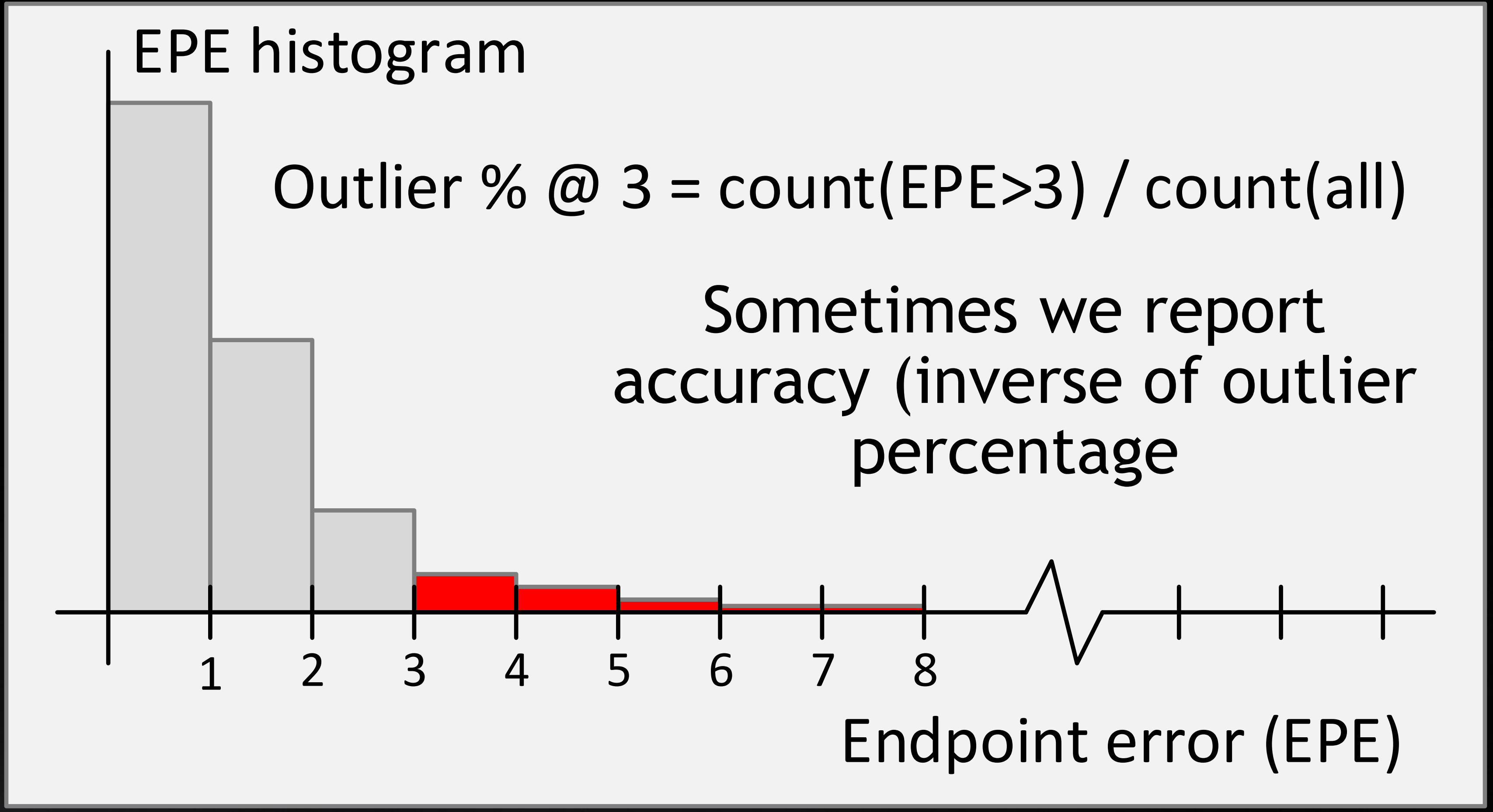


OPTICAL FLOW QUALITY METRIC DEFINITIONS

Endpoint error (EPE)



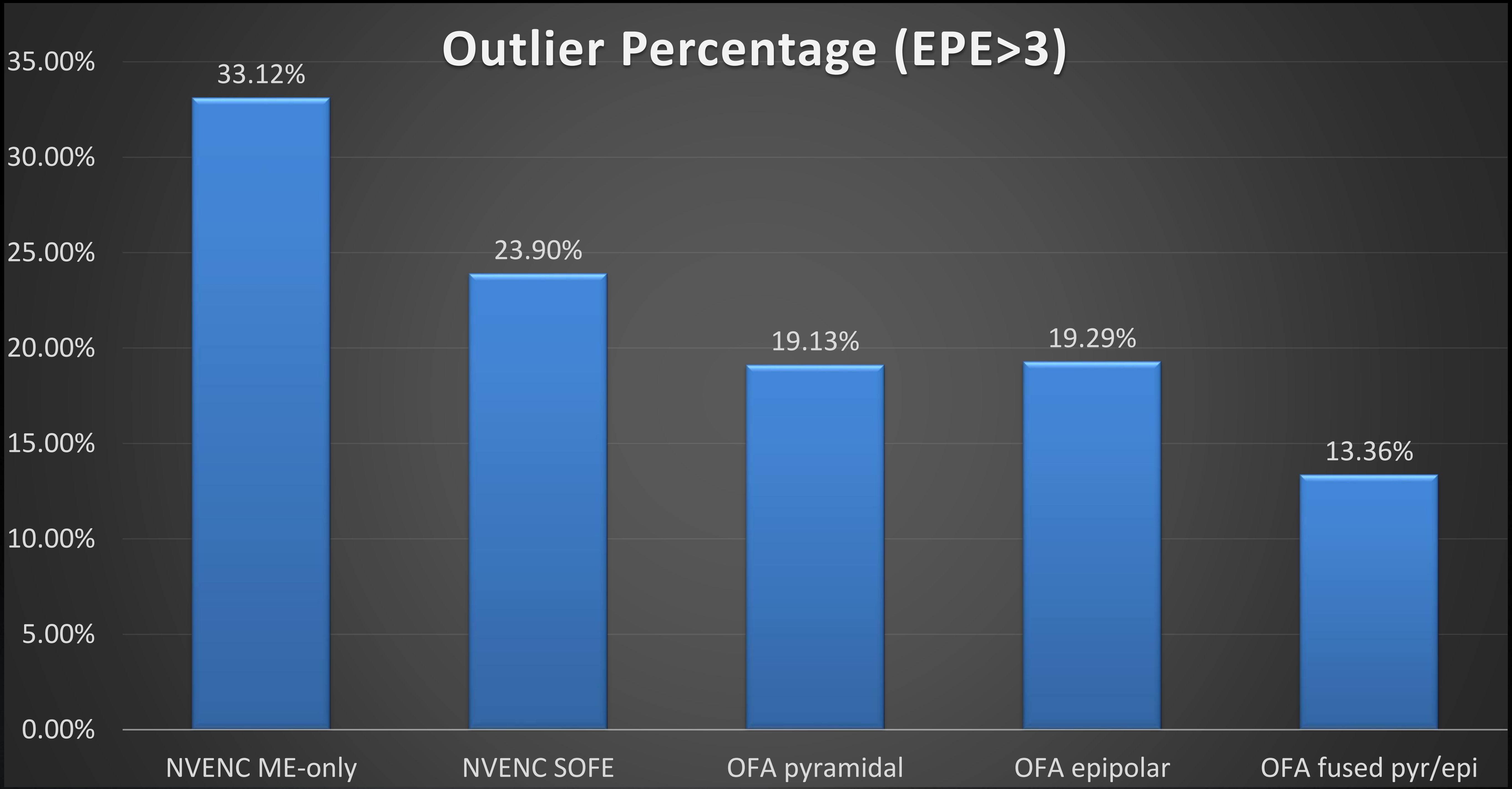
Outlier percentage



We use the same metrics for stereo disparity, although the arrows are always horizontal.

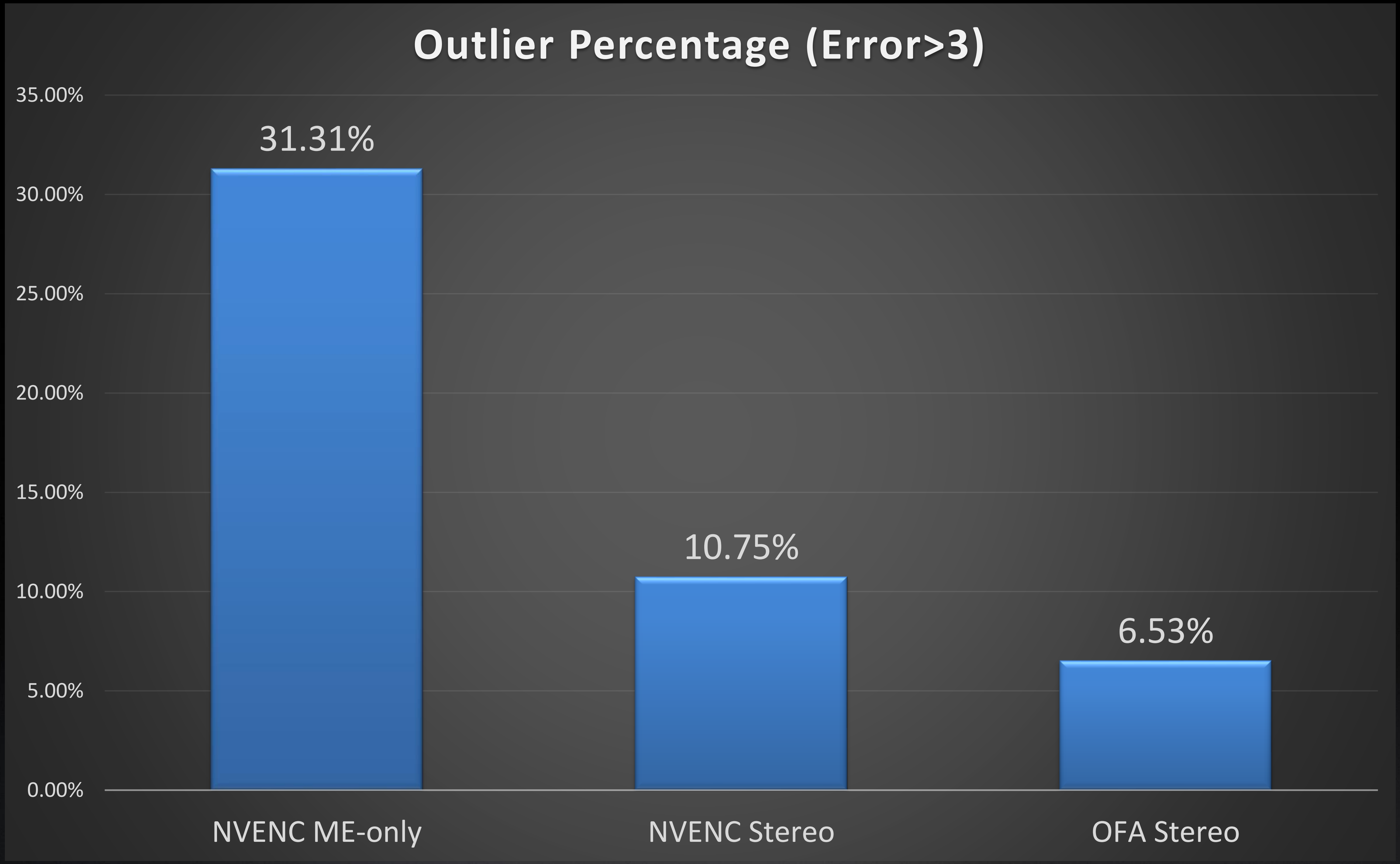
OPTICAL FLOW QUALITY SCORES – HW EVALUATION

KITTI 2015, Optical flow



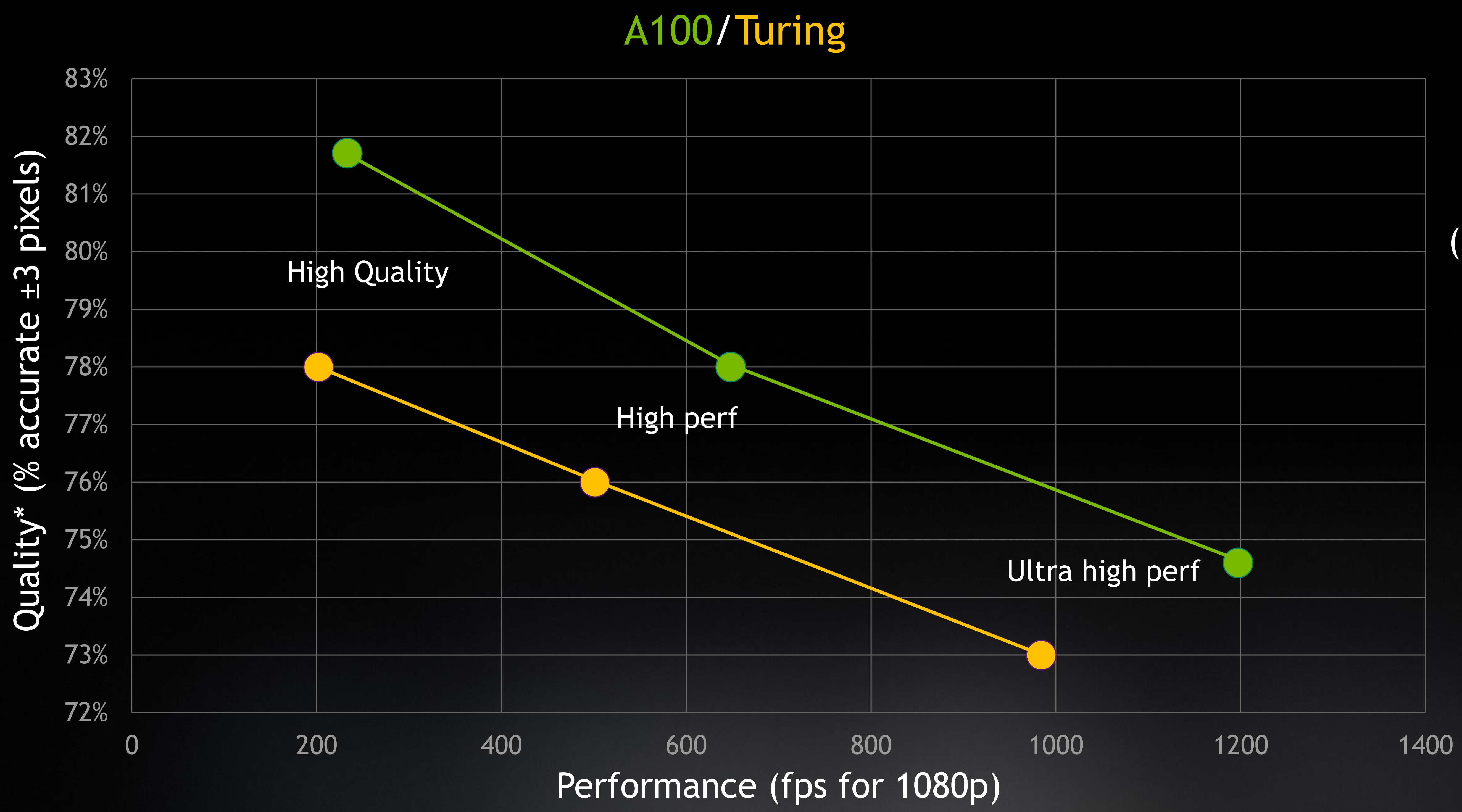
STEREO QUALITY SCORES - HW EVALUATION

KITTI 2015 Stereo



OPTICAL FLOW SDK 2.0

Quality vs Performance



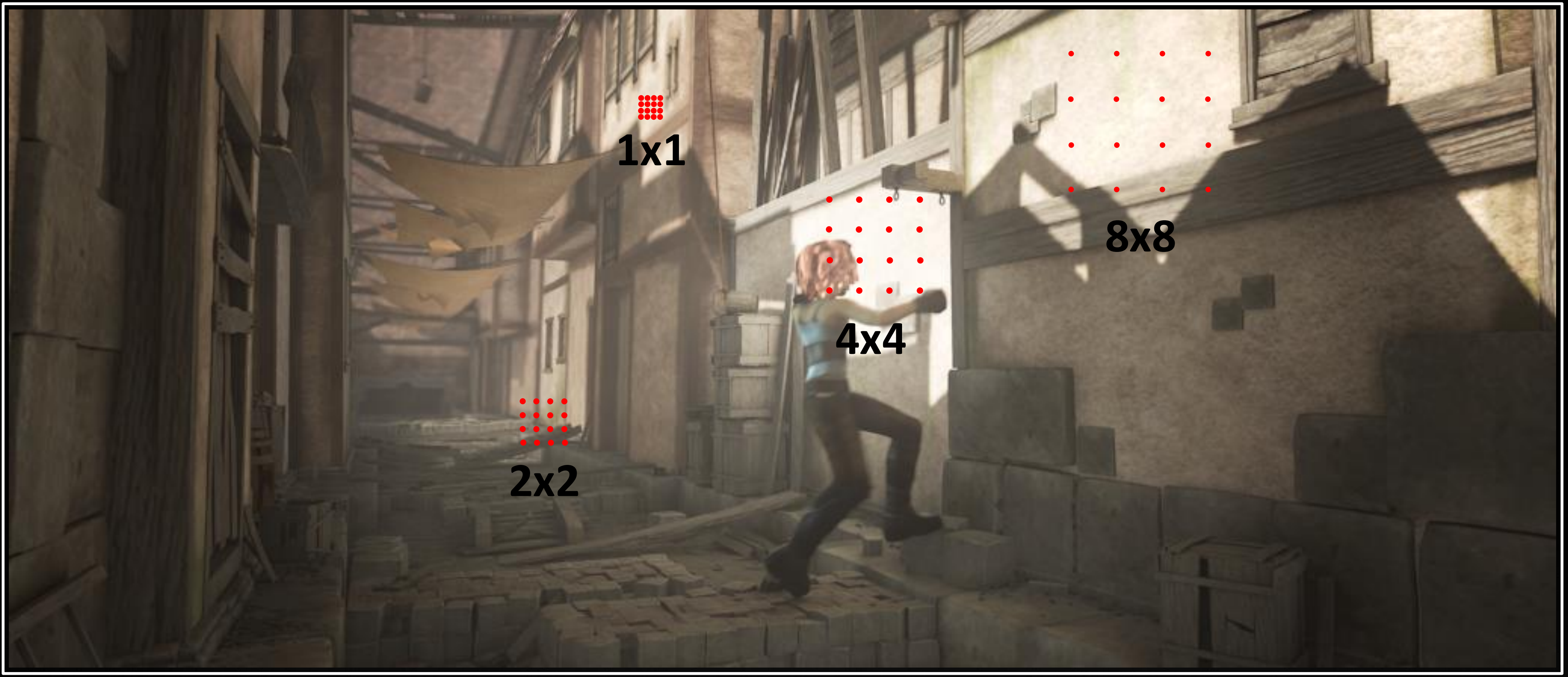
Accuracy vs. speed tradeoff achieved by tuning parameters

(inlier % at difference of 3 pixels)

*KITTI 2015 FL-ALL NOC

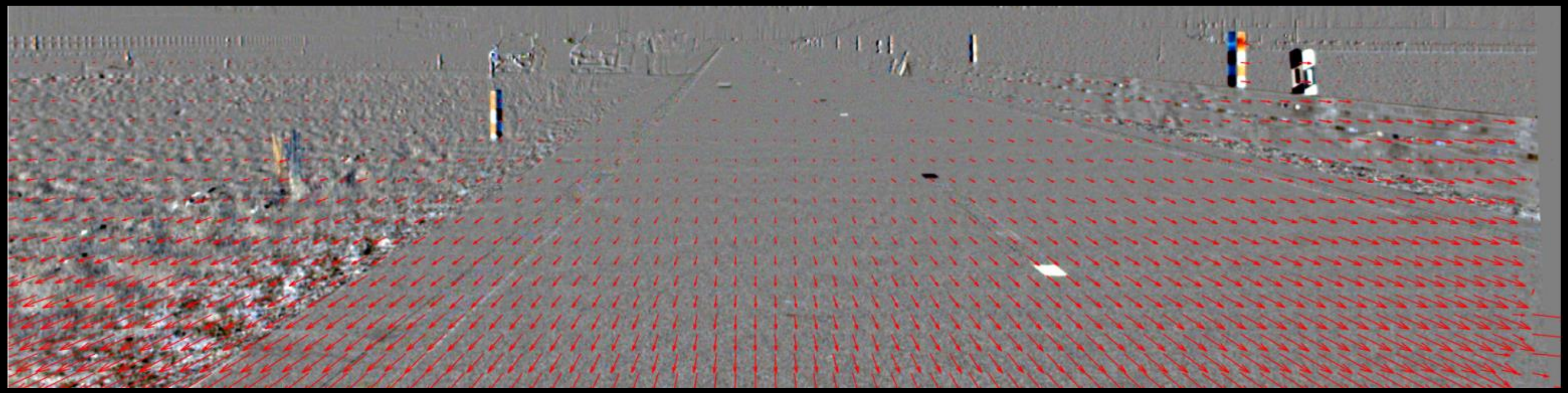
PROGRAMMING FLEXIBILITY

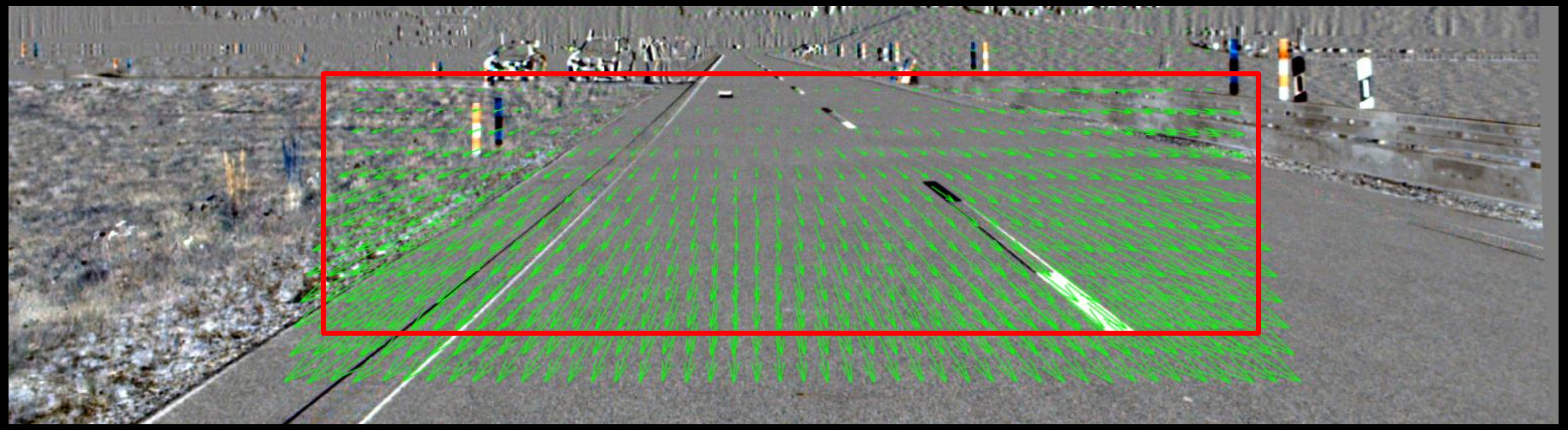
OUTPUT DENSITY



- Select density based on application
- Trade off smaller details for speed
- Use higher density with smaller ROIs

REGION OF INTEREST





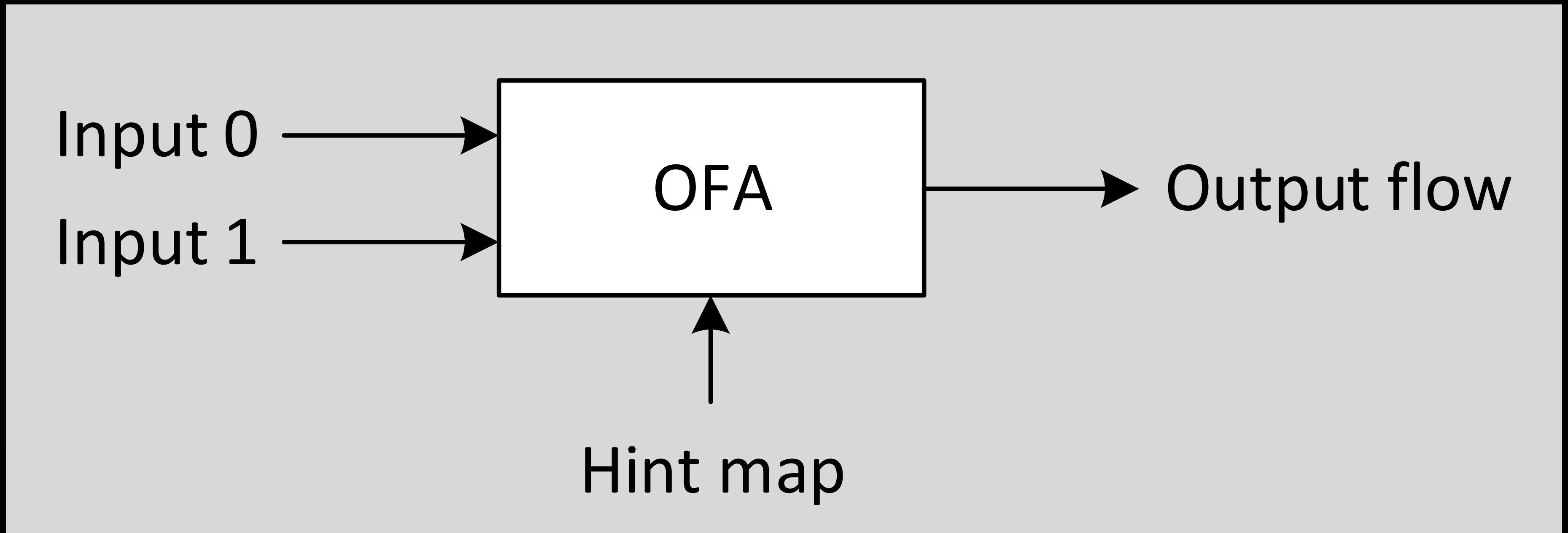
Identify object or area of interest

Define ROI with extended bounds ±*N* pixels

OFA only processes that area

Increase speed by focusing the engine

USING HINT MAPS



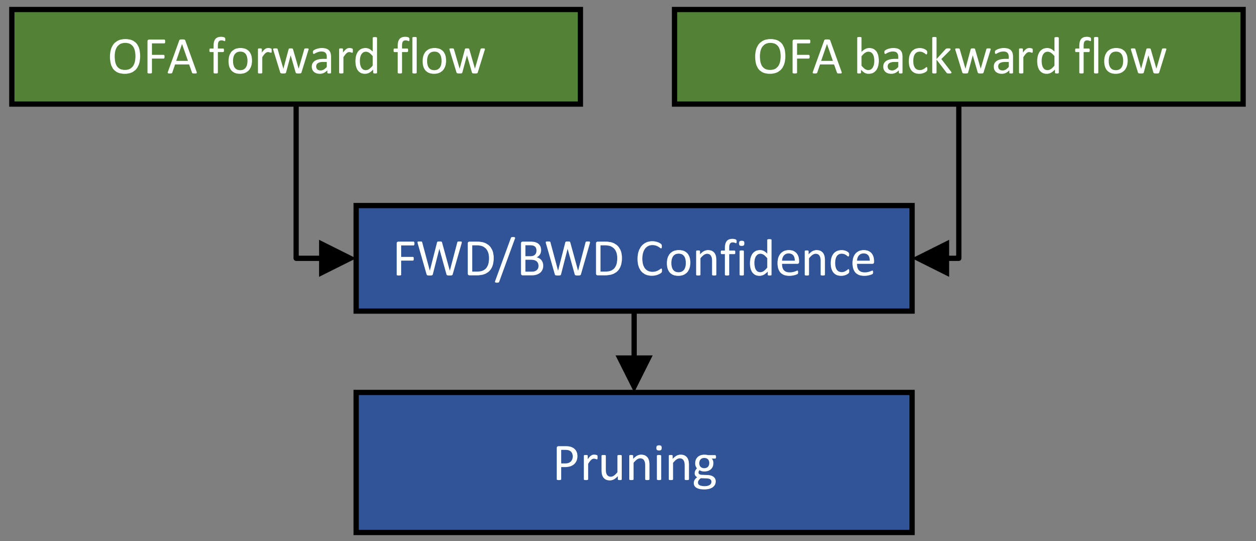
For stereo and pyramidal optical flow

Hint map guides search area

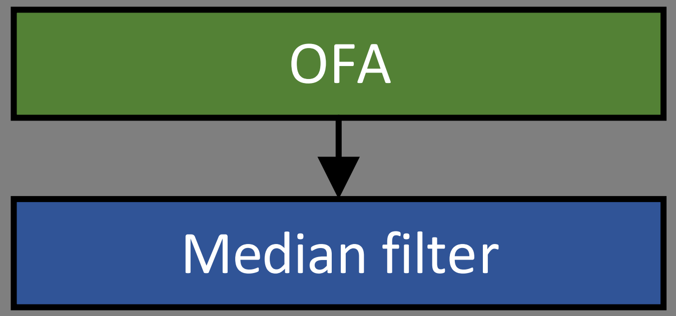
Can come from any source

PROCESSING PIPELINES FOR CONFIDENCE MAPS AND POST-PROCESSING

Confidence generation with forward/backward consistency check

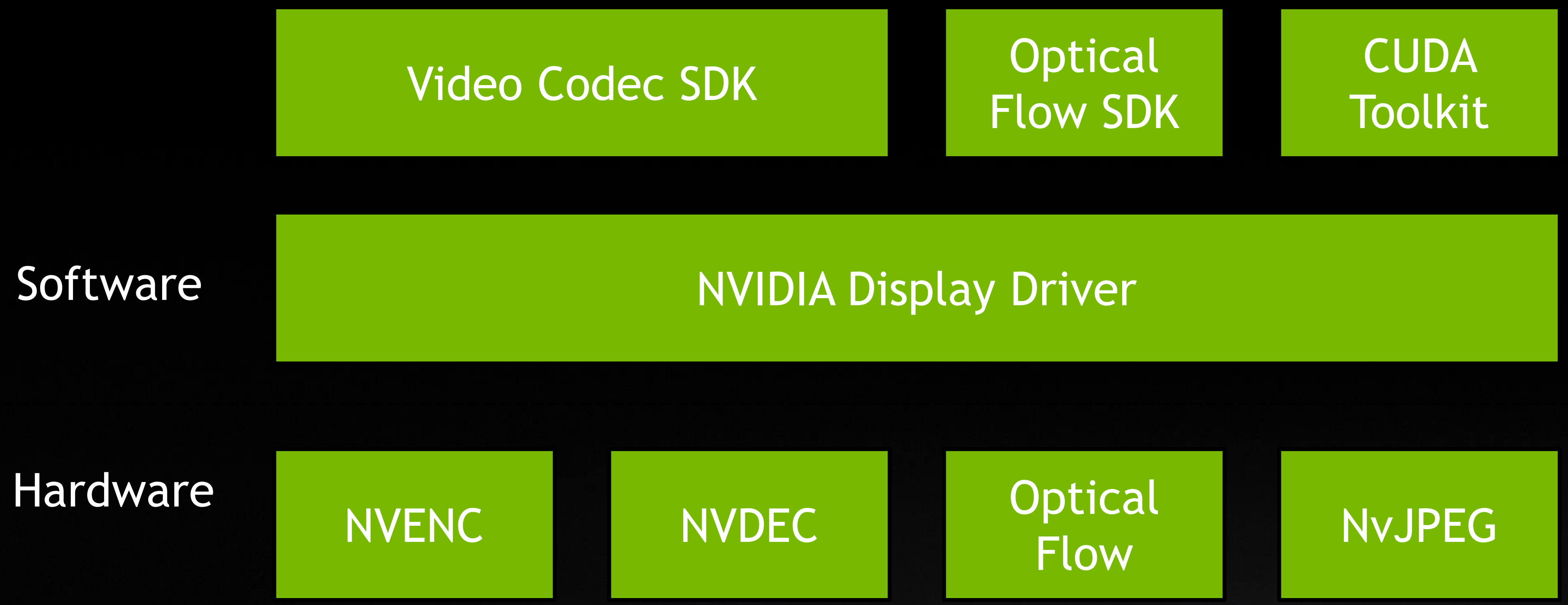


Optical Flow with post-processing for error correction



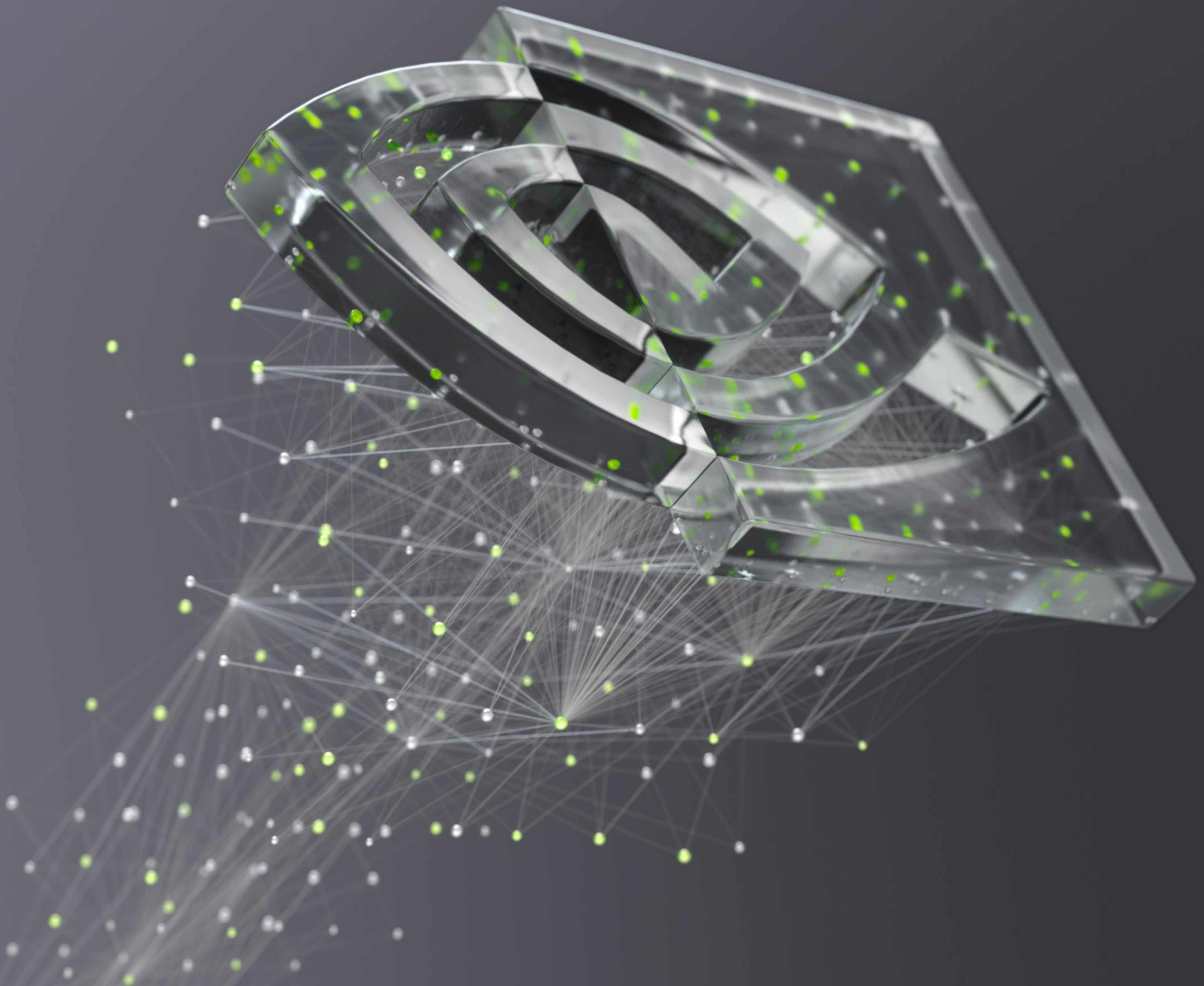
- Confidence maps are often required for downstream processing
- Post-processing can remove outliers
- Done in SW, flexible

SOFTWARE



- All binaries in NVIDIA driver
- SDKs
    - APIs
    - Reusable samples
    - Documentation
- Binary backward compatibility
- Linux & Windows
- CUDA & DirectX APIs

For more details, see: S21337: NVIDIA Video Technologies: Video Codec and Optical Flow SDK



NVIDIA