Progressive Dense Correspondence with Applications to Video Analysis

Mark Duchaineau

Center for Applied Scientific Computing

Lawrence Livermore National Laboratory

CASIS Workshop, Lawrence Livermore National Laboratory

October 17, 2006

We are devising a video browser that can potentially speed human analysis by 100x

- Fully automated video analysis is a long ways off – human analysis will be needed for many years
- Video is world's largest data source
  - VidCharts solution: visually summarize video automatically, provide details interactively
- Computational challenge: design and scaling of correspondence algorithm
  - New GPU-based method promises 100x speedup
- Several national security applications will benefit

Are gigapixel video cameras feasible in the next few years? Yes, but…

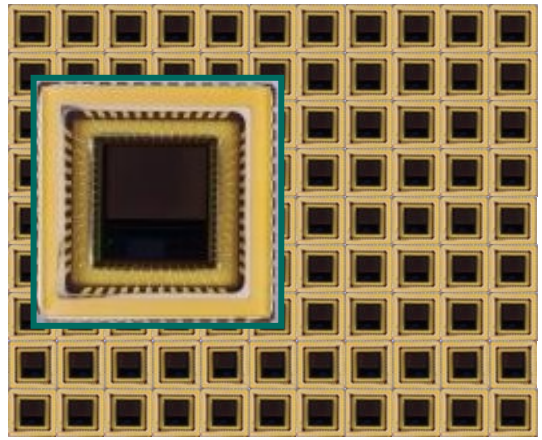
Recipe for Gigapixel Video?:

- 320 KAC-3100 CMOS sensors
  - 2048x1536 2.7um pixels
  - 12Hz full-frame readout, 128mm
  - Noise and packaging issues
- 4 professional photographer lenses
  - Not high enough resolution

  => custom optics with non-planar focus
- Major innovations in packaging, assembly and calibration needed

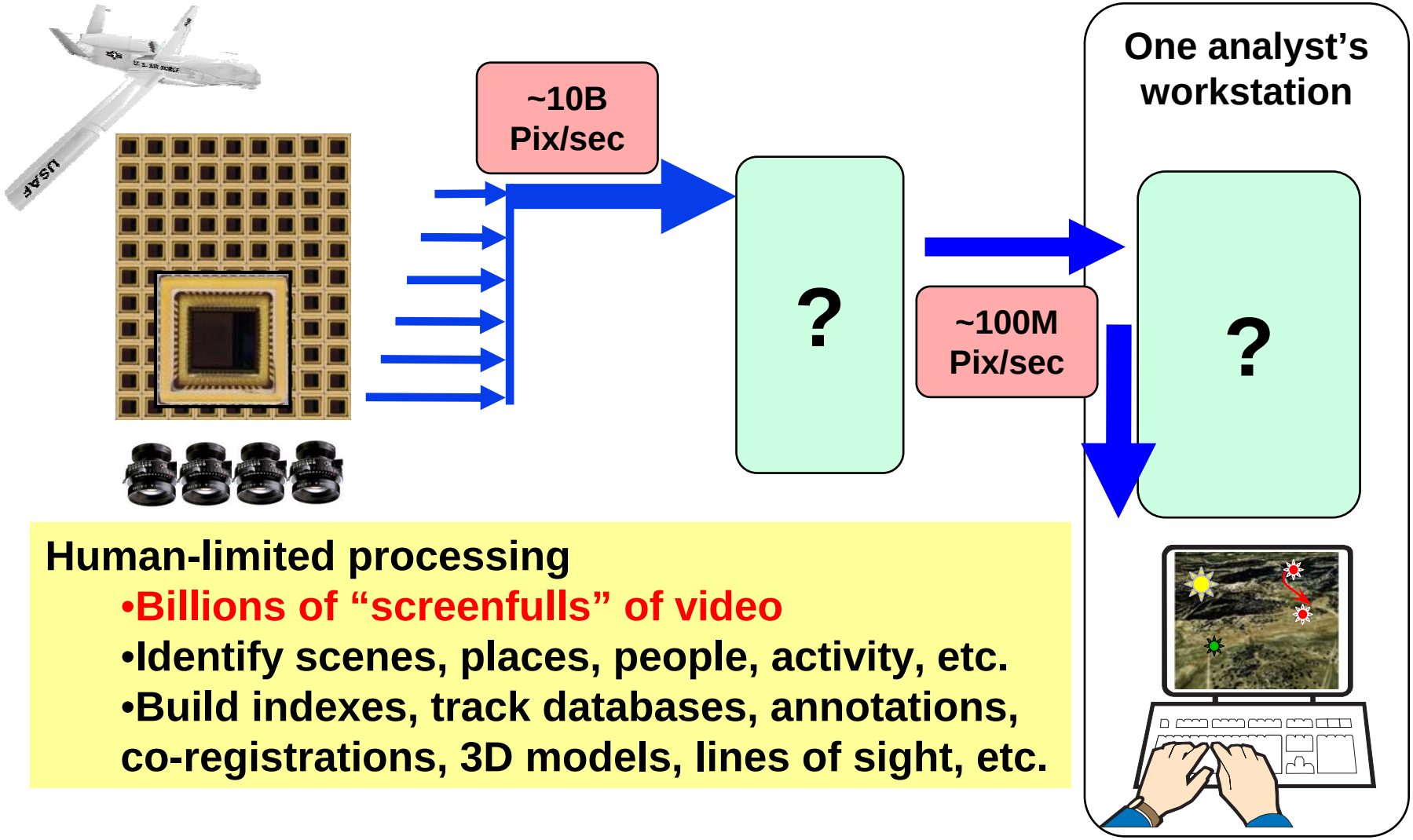
  => MIT LL packaging and automation
- Massive supercomputer onboard

  => Progressive pipeline on 320 FPGAs + 32 GPUs+ 16 IBM Cell chips



- Pros
  - 32000x32000 pixels
  - Catches transitory events
  - Many off-the-shelf parts
- Cons
  - 1.0 terabyte per minute
  - Massive stream compute must be near the sensor

Challenge: assist human analysis of massive collections of large-scale video frames



Human-limited processing

- **Billions of "screenfulls" of video**
- **Identify scenes, places, people, activity, etc.**
- **Build indexes, track databases, annotations, co-registrations, 3D models, lines of sight, etc.**

Realtime Optimally Adapting Mesh (ROAM) processing optimizes sensor-to-user data-flow with selective refinement

GPU-accelerated, progressive, airborne analysis pipeline

Raw Sensor Image 10-100M Pix/sec

10-100M Pix/sec

10-100M Pix/sec

10-100M Pix/sec

10-100M Pix/sec

10-100M Pix/sec

Build hierarchy

hierarchy

hierarchy

hierarchy

hierarchy

hierarchy

Frame-to-Frame warp

Video Lifting of layers

Motion Analysis

1M-1B Pix/sec

Wavelet Time+space Compress

100K-100M Bits/sec

One analyst's workstation

Decompress

Feature Extraction

Analysis

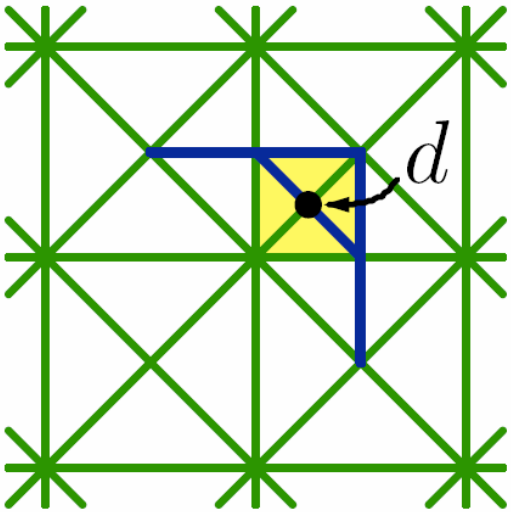
Visualization

ROAM streams are scalable, matched to available resources

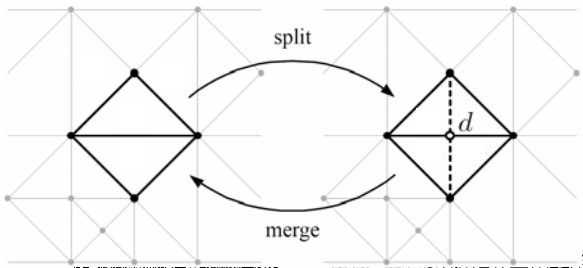
- **End-to-end optimality feedback**
- **Real-time analysis**

ROAM is our selective-refinement, tiled-stream processing foundation

Uses 4-8 meshes with diamond data structure

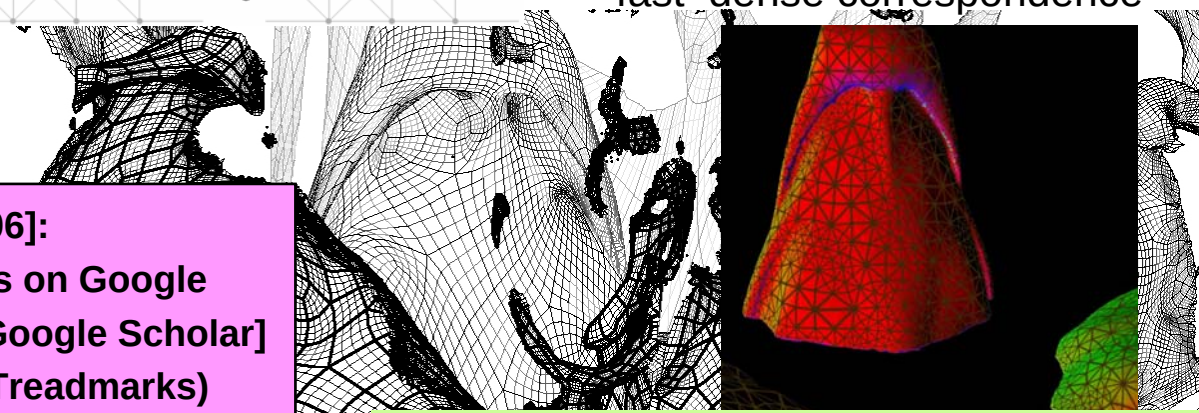


Dual-queue incremental updates exploit frame-to-frame coherence



Flexible framework – many optimizations, extensions

- frustum cull
- defer priority compute
- triangle patches
- texture tiles
- line-of-sight
- 3D diamonds
- fast dense correspondence

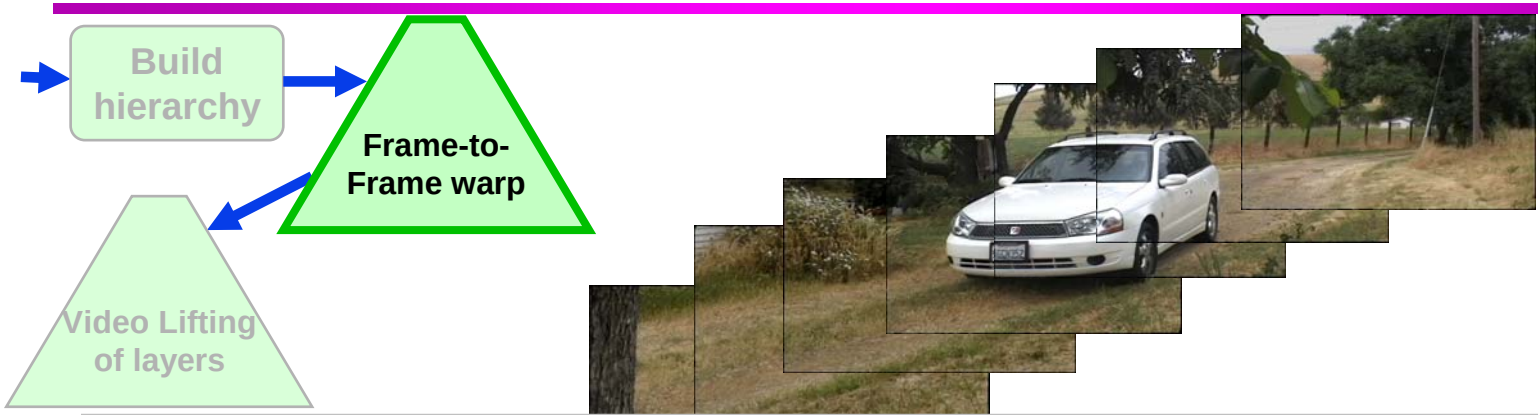


Claims to fame [As of Sept 25, 2006]:

- **"ROAM Algorithm" has 1,150 hits on Google**
- **'97 paper has 381 citations [via Google Scholar]**
- **Used in several game titles (e.g. Treadmarks)**
- **Los Alamos "Outstanding Innovation" award**
- **Vis '04 "Best Paper" (Hwa/Duchaineau/Joy)**

ROAMing a shrink-wrapped isocontour from an 8B elem Gordon-Bell Prize run

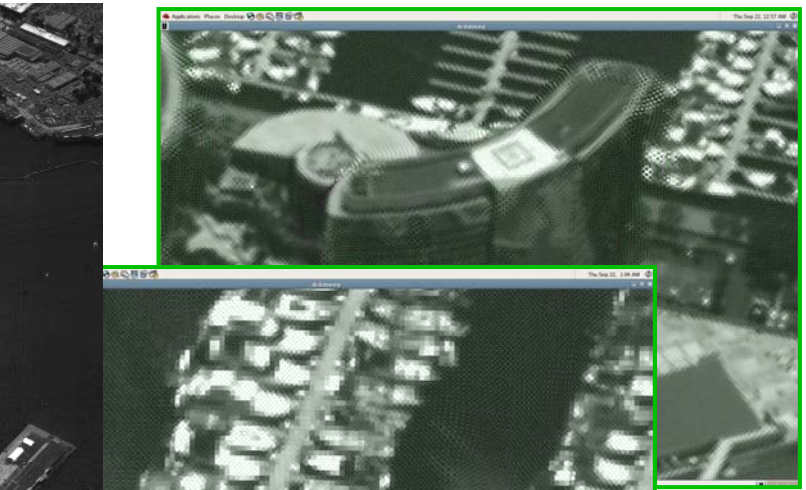
Image warp will enable a video summary and drill-down system



- Goal: browse hours of video information in minutes
- Warping computes dense frame-to-frame correspondence
- Lifting summarizes what is in common over a long sequence
- Sparse mismatch = mover, large mismatch = scene change
- Algorithm tracks many kinds of motion:
  - Pan, zoom, rotation
  - Track rigid and deformable movers against background
  - Detect and correct for brightness changes

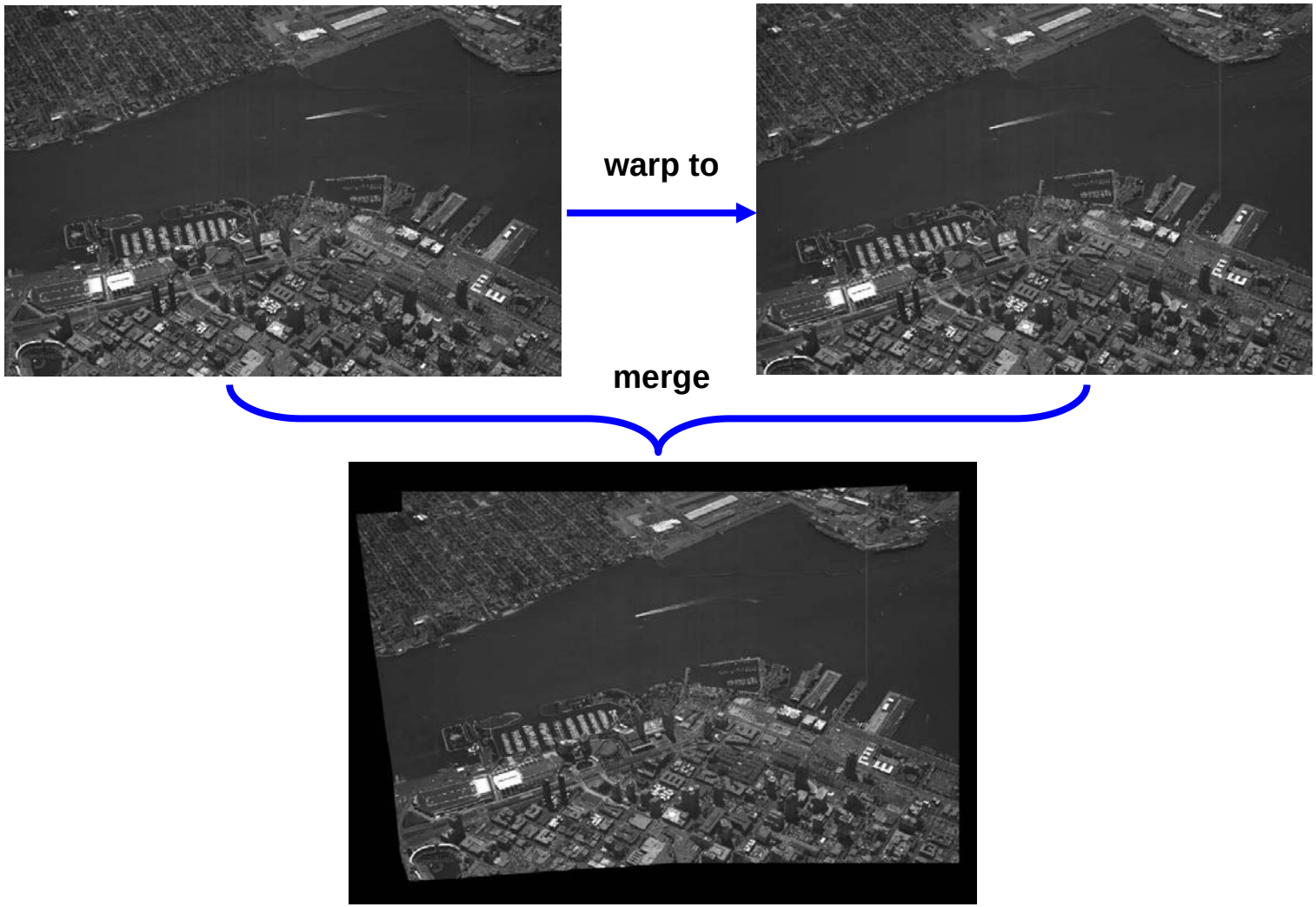
We develop new, ROAM-based dense image correspondence algorithms

- Proof-of-concept test on aerial video
- Super-resolution effect implies sub-pixel accuracy
- Potential for improved processing speed, compression
  - CPU version not fast enough – GPU port of kernels is fast
- Possible use for 3D from video
- Current acceleration: 7.4 million pixels per second registered per GPU

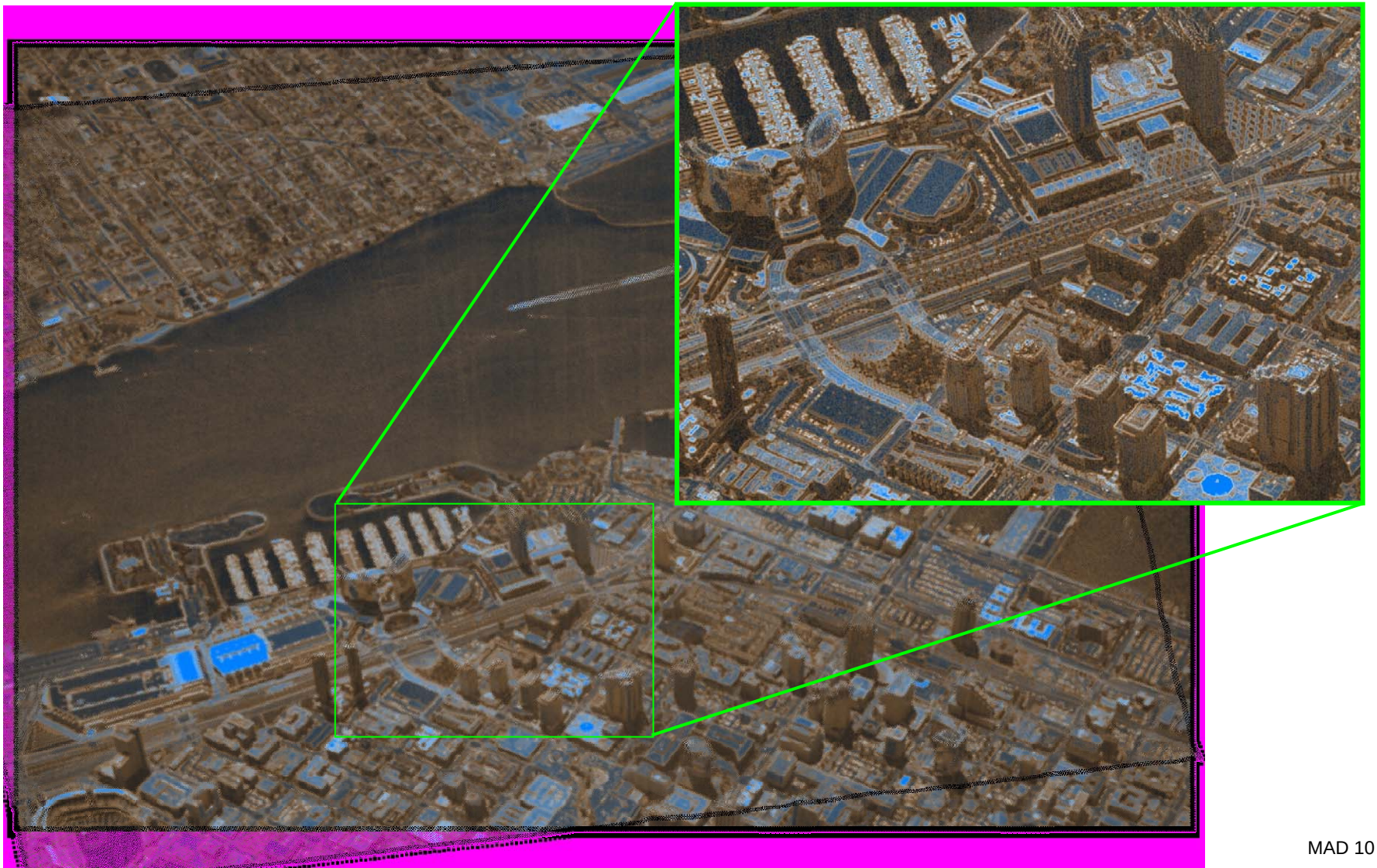


Correspondence is computed for pairs of images from the video sequence

- **Frames 10 and 31 from aerial video sequence:**

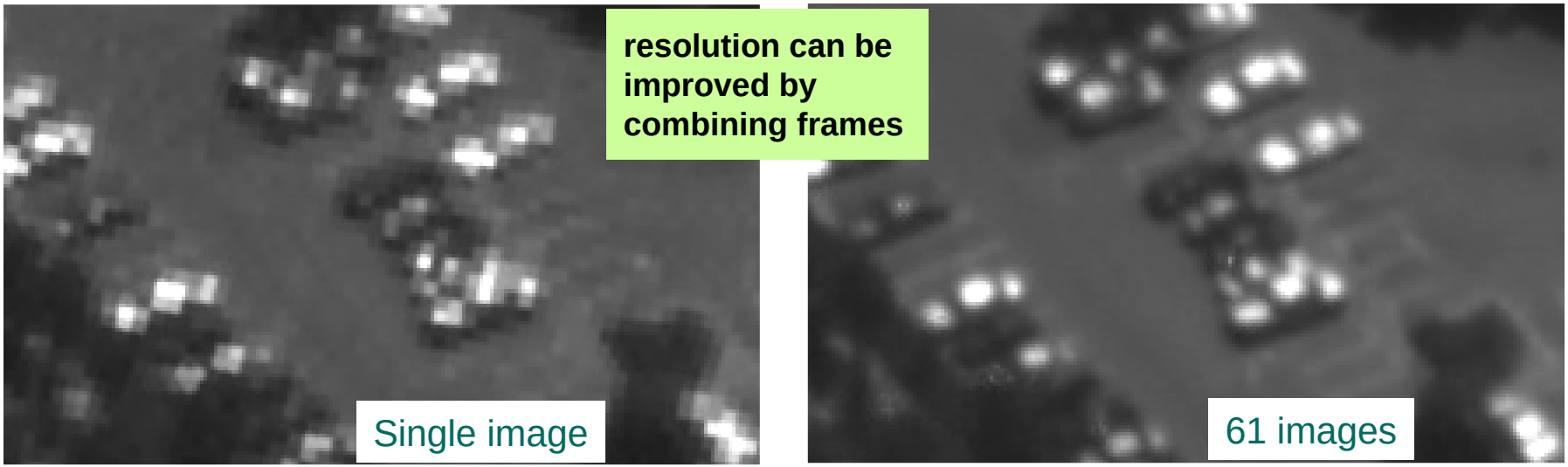


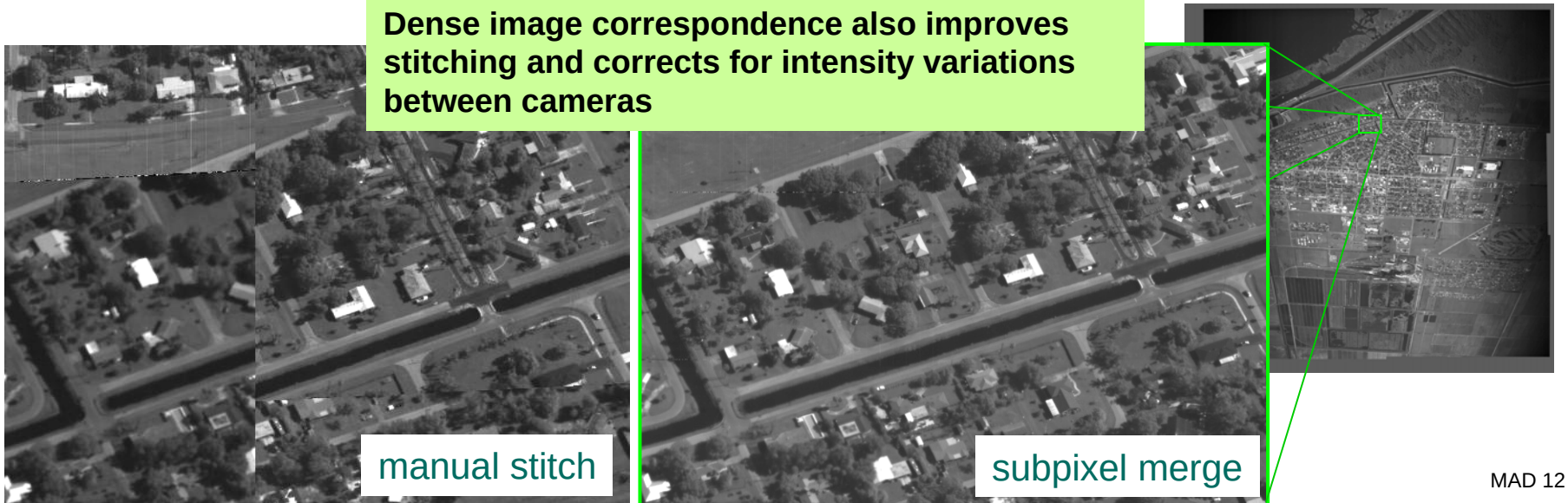
Selective refinement of dense correspondence is robust, maps well to GPU acceleration



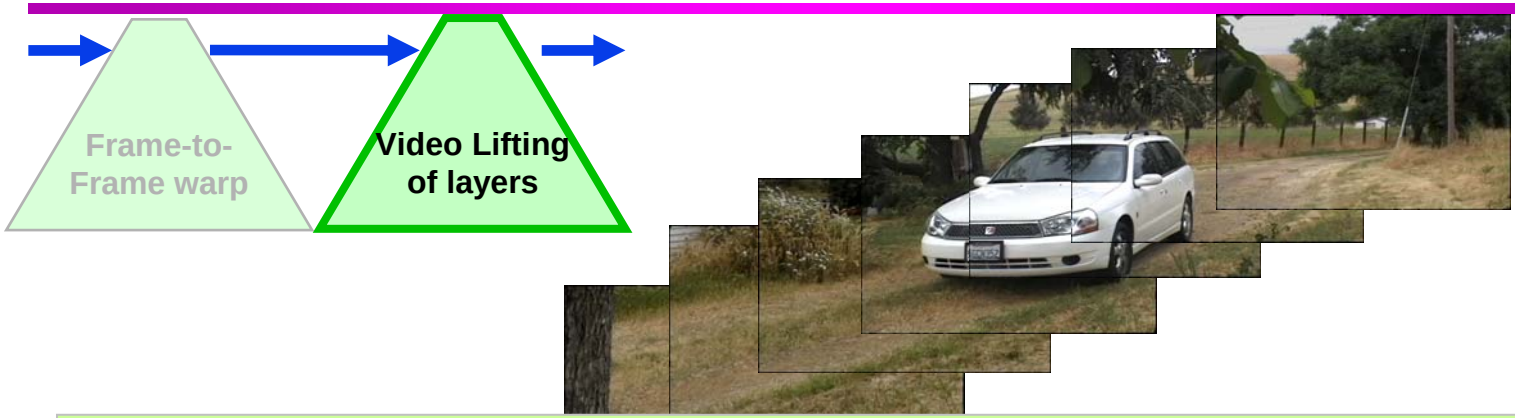
Correspondence algorithm successfully stabilizes the imagery

Dense correspondence enables resolution improvements and seamless mosaic stitching



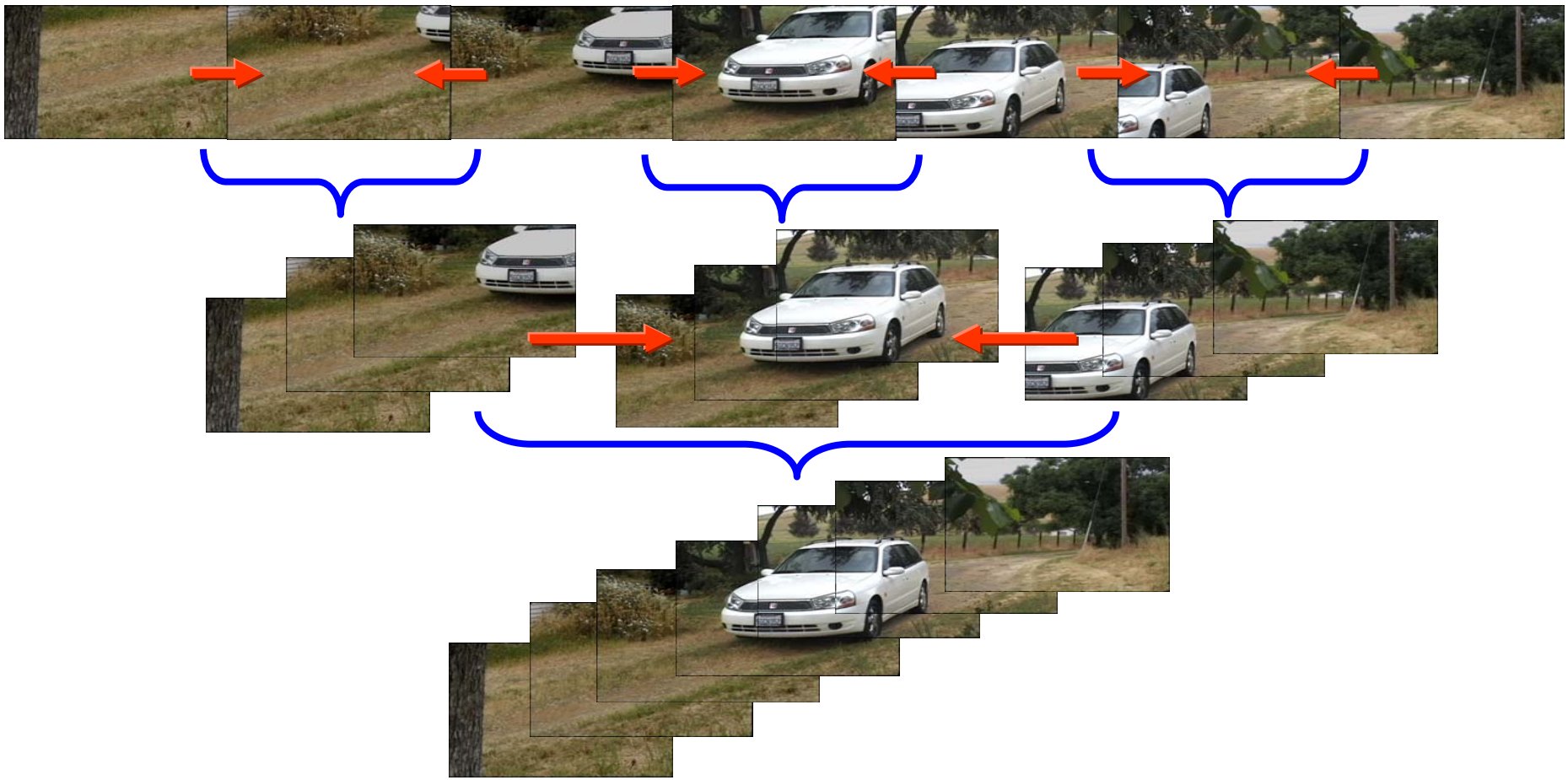


Video “lifting” extends the correspondence to video sequences



- Goal: browse hours of video information in minutes
- Warping computes dense frame-to-frame correspondence
- **Lifting summarizes what is in common over a long sequence**
- Sparse mismatch = mover, large mismatch = scene change
- Algorithm will track many kinds of motion:
  - Pan, zoom, rotation
  - Track rigid and deformable movers against background
  - Detect and correct for brightness changes

Video “lifting” extends the correspondence to video sequences

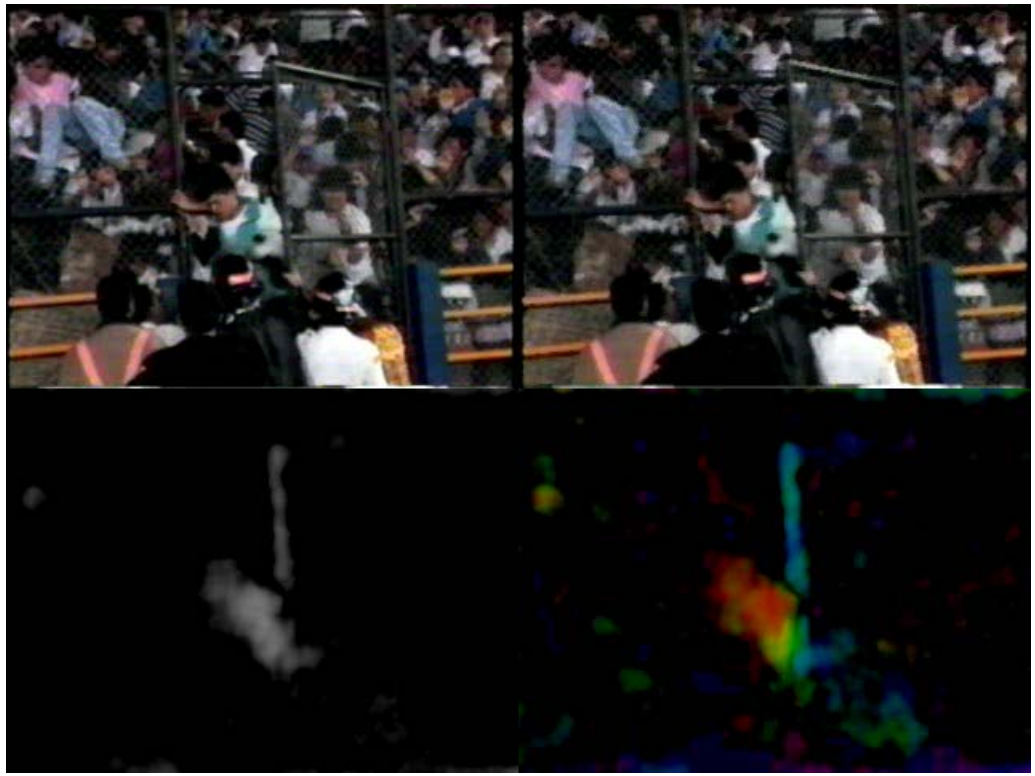


Multi-scale progressive warping stabilizes imagery and facilitates automated detection of movers

m

Image warp appears to work well on complex video sequences

- Deformable moving objects
- Complex lighting and occlusion
- Erratic camera motion
- Foreground/background issues
- Motion blur



VidCharts will provide orders of magnitude improvements in video analysis

VidCharts will:

- **Provide a 100x speedup in human analysis**
- **Speed video analysis computation 100x**
- **Allow in-situ analysis (near sensors)**
  - **Avoid bottlenecks to central facility**
- **Develop new video summary and drill-down capabilities**