Computational Photography and Capture:

Video Based Rendering

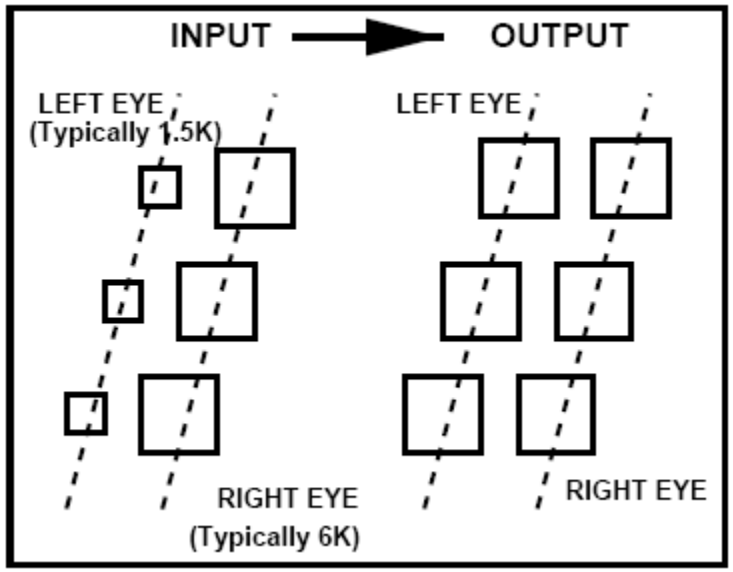
Gabriel Brostow & Tim Weyrich

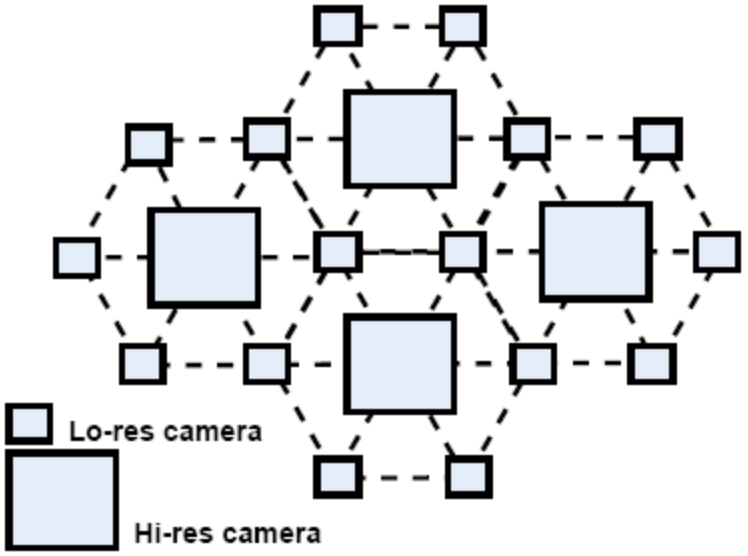
TA: Frederic Besse

Hybrid Stereo Camera: An IBR Approach for Synthesis of Very High Resolution Stereoscopic Image Sequences

Sawhney, Guo, Hanna, Kumar, Adkins, Zhou, **SIGGRAPH 2001**

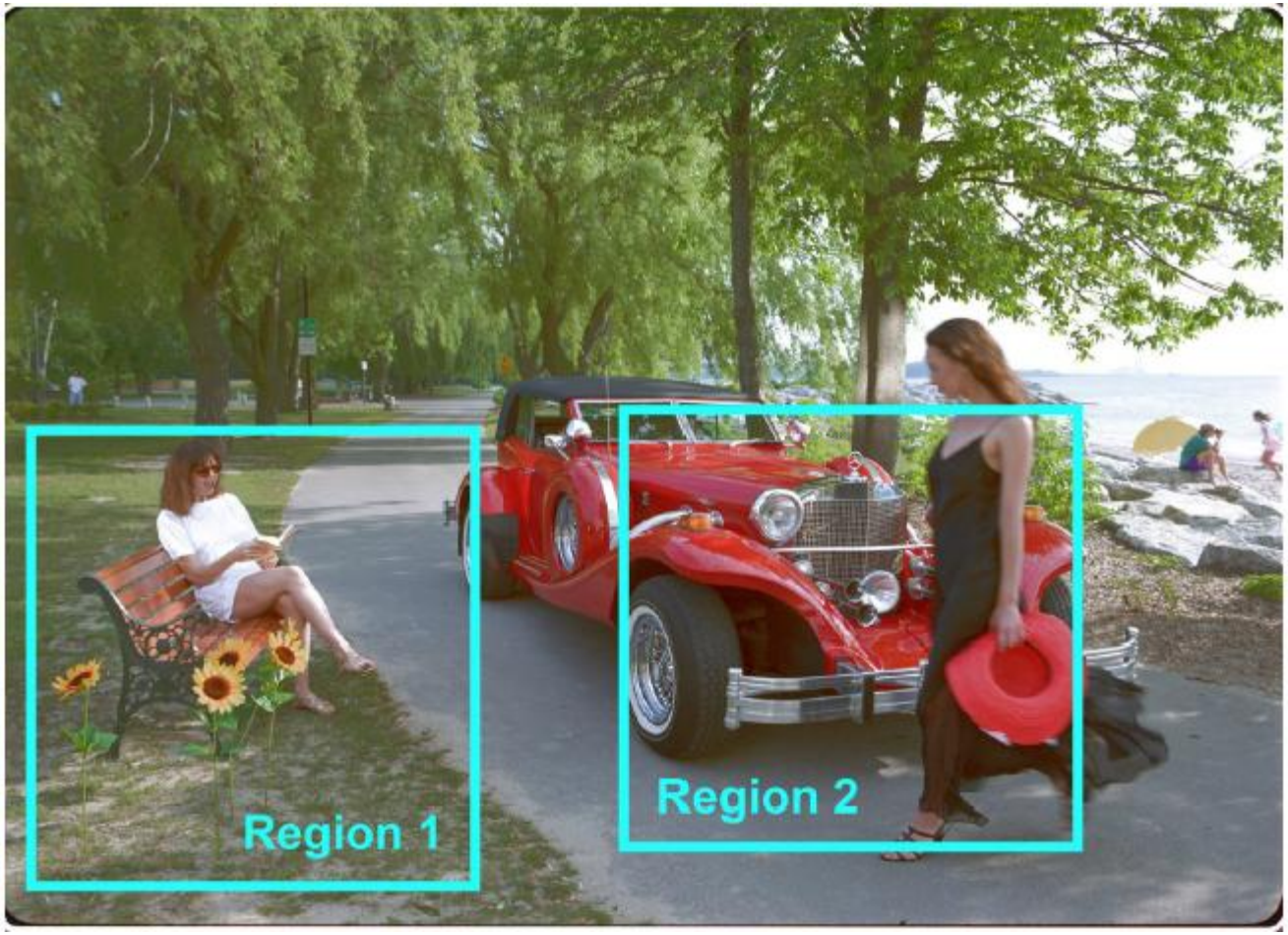
- Stereo films are great but cameras are heavy
- High resolutions require years of rendering
  - "100 computers for 1 year for 45-minute film"
- Let us replace stereo-cameras with:





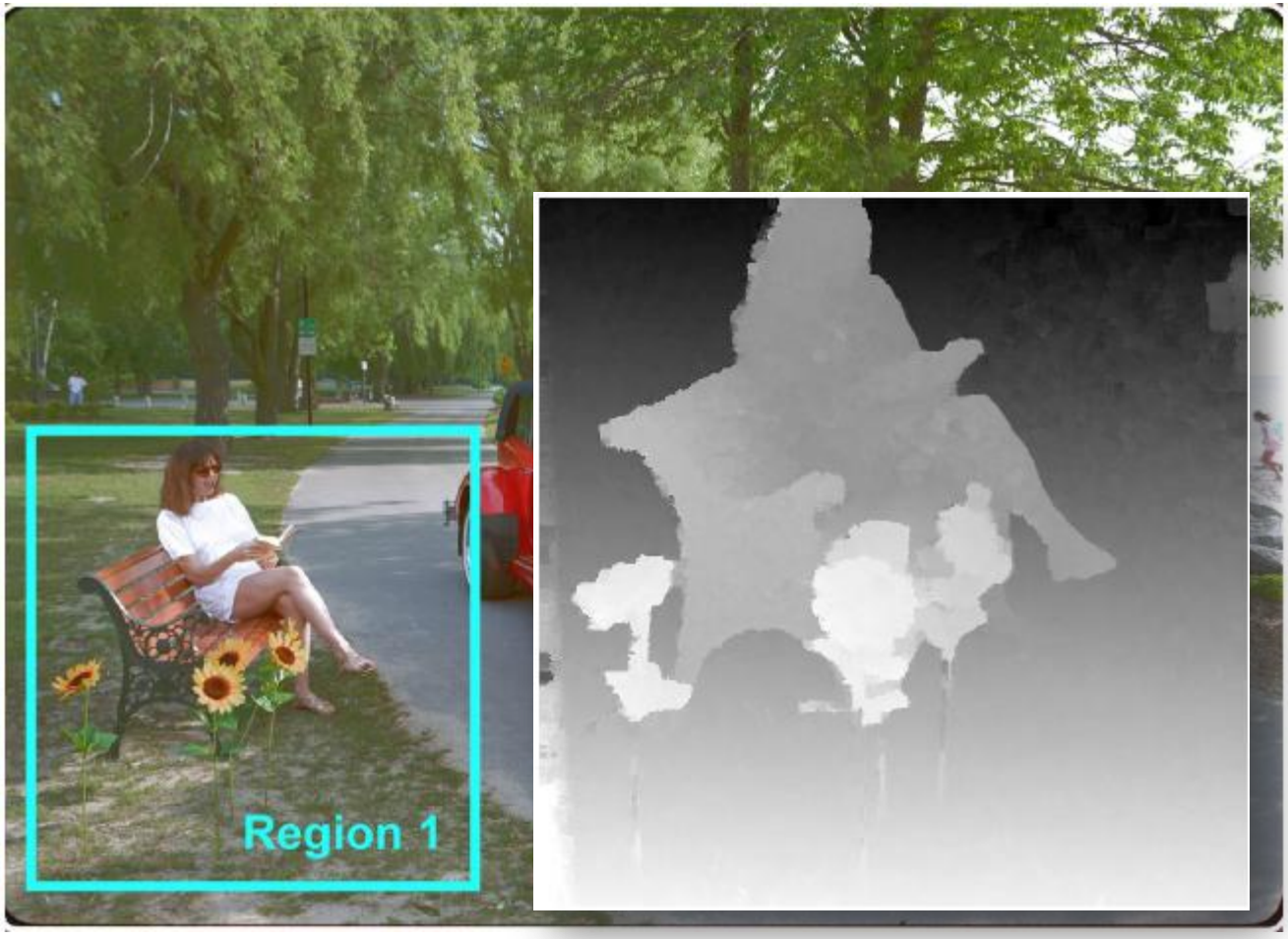
Hybrid Stereo Camera: An IBR Approach for Synthesis of Very High Resolution Stereoscopic Image Sequences

Sawhney, Guo, Hanna, Kumar, Adkins, Zhou, **SIGGRAPH 2001**



Hybrid Stereo Camera: An IBR Approach for Synthesis of Very High Resolution Stereoscopic Image Sequences

Sawhney, Guo, Hanna, Kumar, Adkins, Zhou, **SIGGRAPH 2001**



(A1) Input: Right Original Full-res ($2K \times 2K$)

(A3) Input: Left Low-resolution (512×512)

(A2) Output: Left Synthesized Full-res ($2K \times 2K$)

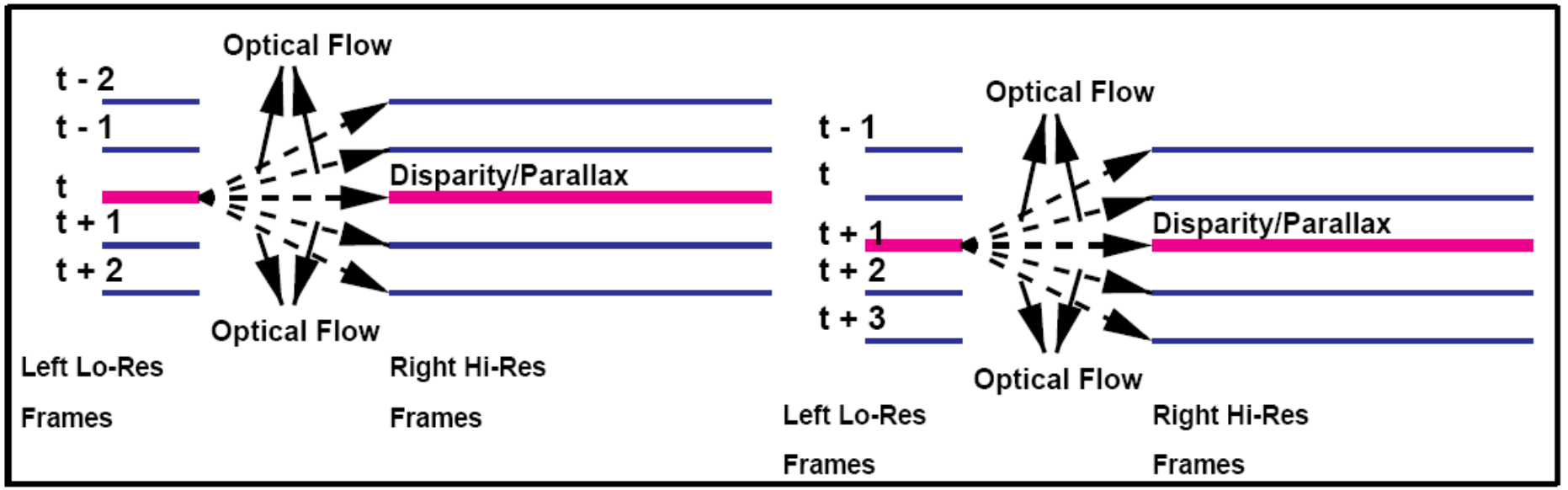
(A1) Input: Right Original Full-res ($2K \times 2K$)

(A3) Input: Left Low-resolution (512×512)

Final Result

Asymmetric stereo perception:
"Frequency components that are not identical will cause binocular rivalry." The high-resolution view will dominate.

Bicubic upsampling of low-resolution original



- Compute <u>per-pixel</u> disparity at t
- Compute flow-corrected disparity at +/- Δt
- Synthesize and evaluate <u>quality measure</u>
  - Finds missing and low-quality pixels
- Fill in those gaps with regular up-sampled pixels

Stereo only

Stereo + Flow

Stereo + Flow + Fillin
(+ Color Correction)

Challenges of Hybrid Stereo

- Depth estimates were only good locally
- Motion tracking is very limited
- (Need at least 1 high-resolution video)

Note:
Using Pravin Bhat's slides with permission

Using Photographs to Enhance Videos of a Static Scene

Pravin Bhat, C. Lawrence Zitnick, Noah Snavely, Aseem Agarwala, Maneesh Agrawala, Michael Cohen, Brian Curless, Sing Bing Kang

EGSR 2007

Project web page

Motivation

Consumer video

Motivation

Consumer Photographs

Motivation

Consumer Photographs

Motivation

Consumer Photographs

Motivation

Consumer Photographs

Motivation

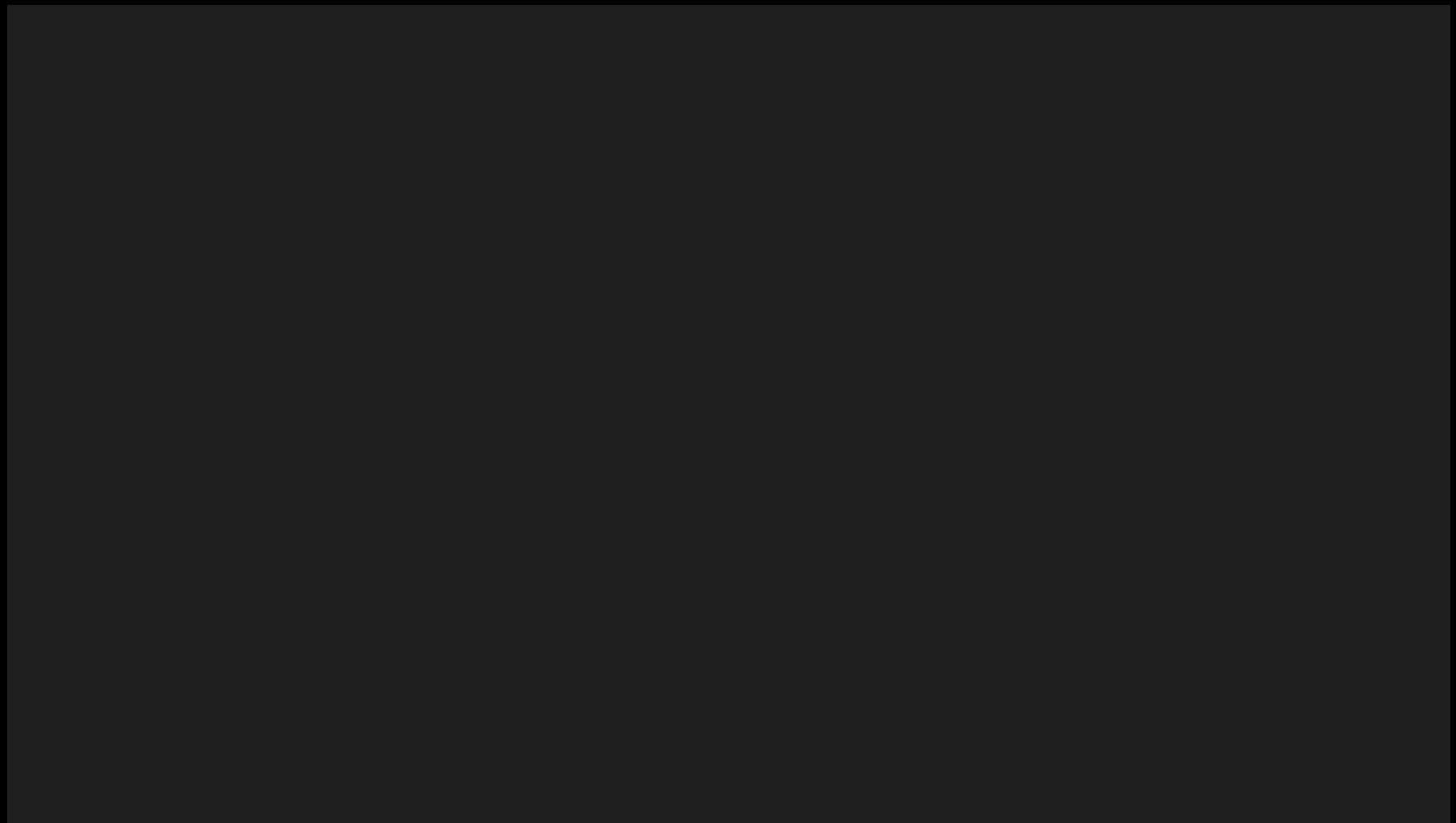
Consumer Photographs

Our approach

Video Enhancements

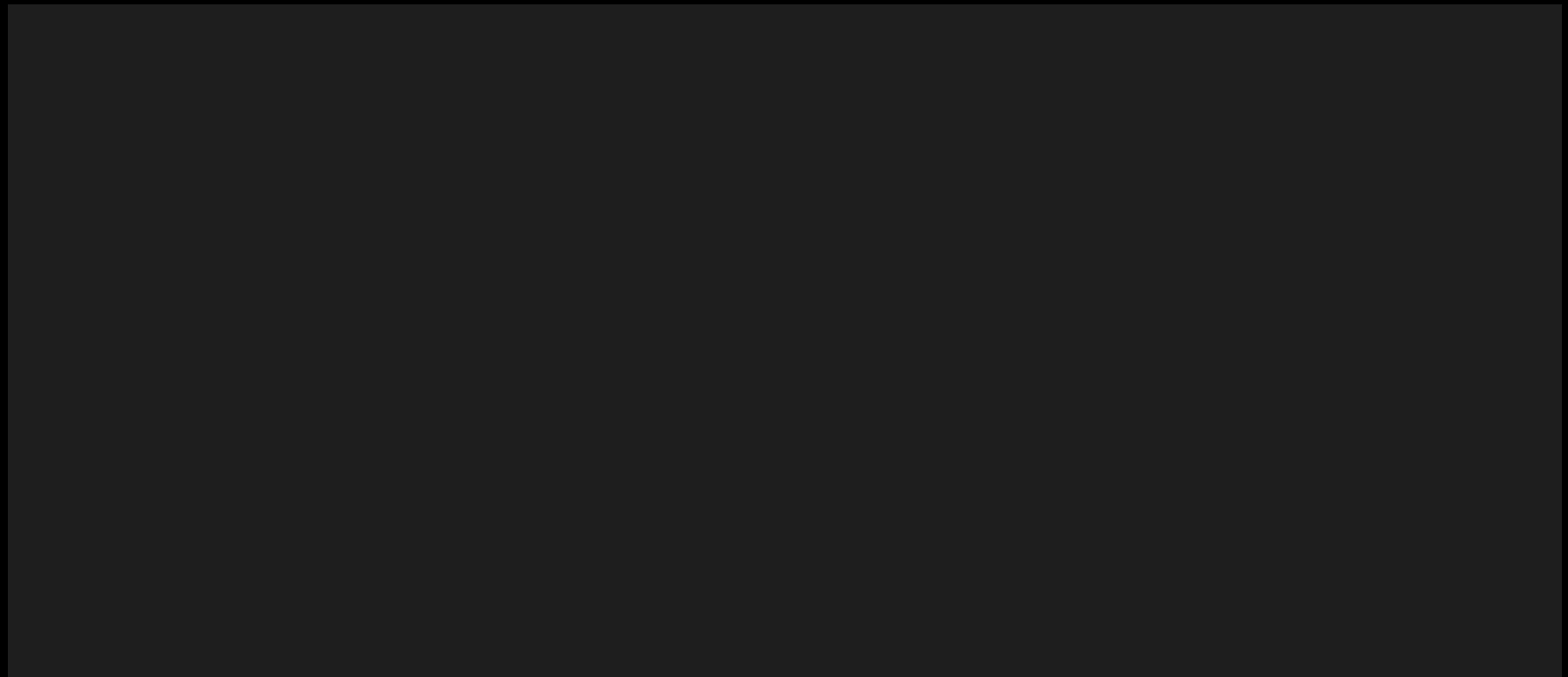
Video Enhancements

- One framework. Variety of enhancements:



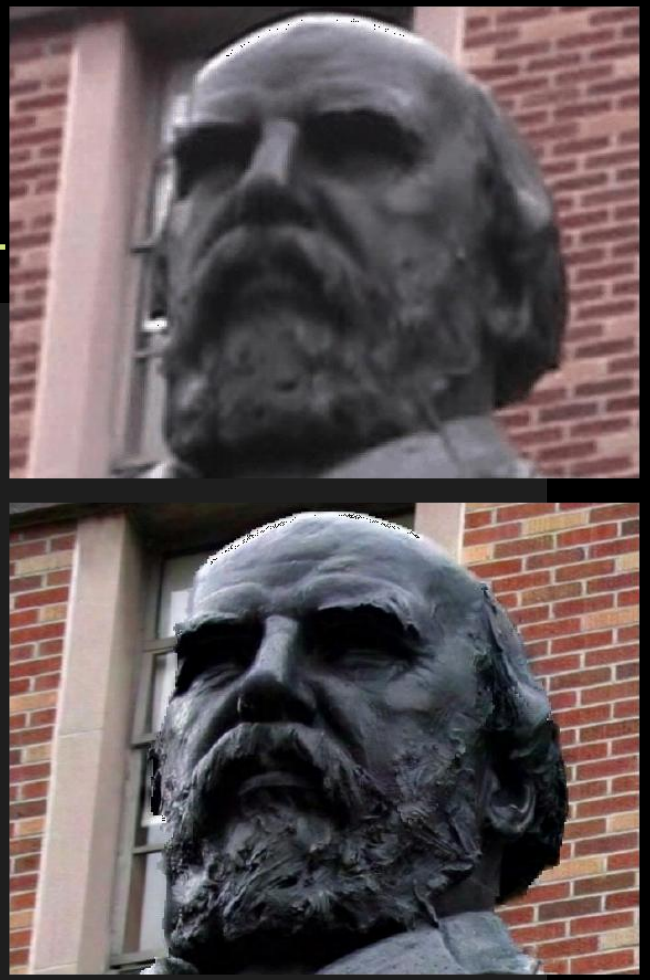
Video Enhancements

- One framework. Variety of enhancements:
  - Transferring photographic qualities



Video Enhancements

- One framework. Variety of enhancements:
  - Transferring photographic qualities
    - resolution



Video Enhancements

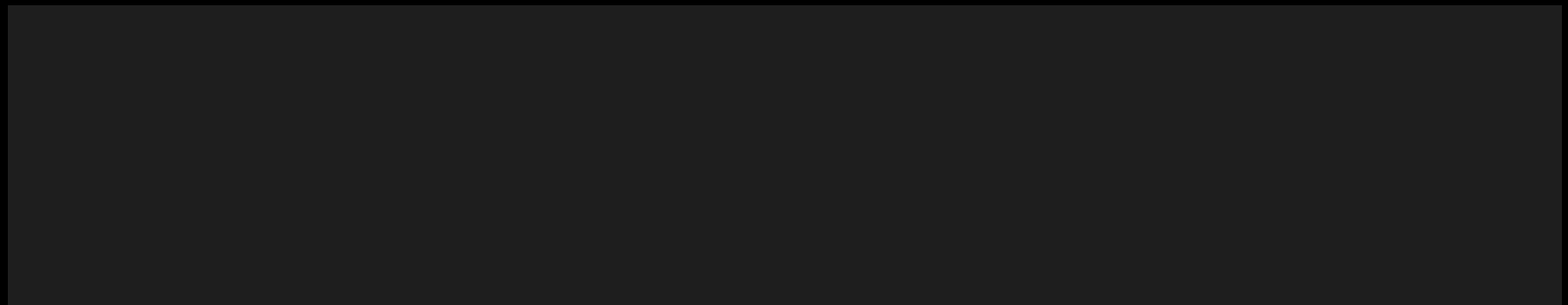
- One framework. Variety of enhancements:
  - Transferring photographic qualities
    - resolution
    - exposure

Video Enhancements

- One framework. Variety of enhancements:
  - Transferring photographic qualities
    - resolution
    - exposure
    - dynamic range

Video Enhancements

- One framework. Variety of enhancements:
  - Transferring photographic qualities
    - resolution
    - exposure
    - dynamic range
  - Easy video authoring



Video Enhancements

- One framework. Variety of enhancements:
  - Transferring photographic qualities
    - resolution
    - exposure
    - dynamic range
  - Easy video authoring
    - object touch-up

Video Enhancements

- One framework. Variety of enhancements:
  - Transferring photographic qualities
    - resolution
    - exposure
    - dynamic range
  - Easy video authoring
    - object touch-up
    - image filters

Video Enhancements

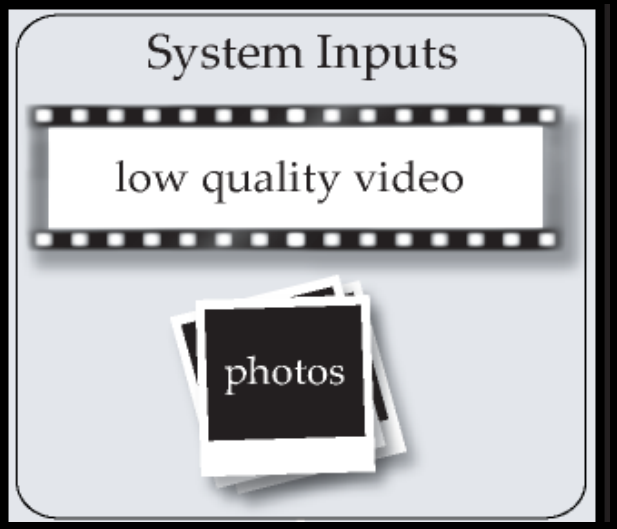
- One framework. Variety of enhancements:
  - Transferring photographic qualities
    - resolution
    - exposure
    - dynamic range
  - Easy video authoring
    - object touch-up
    - image filters
    - object removal

Video Enhancements

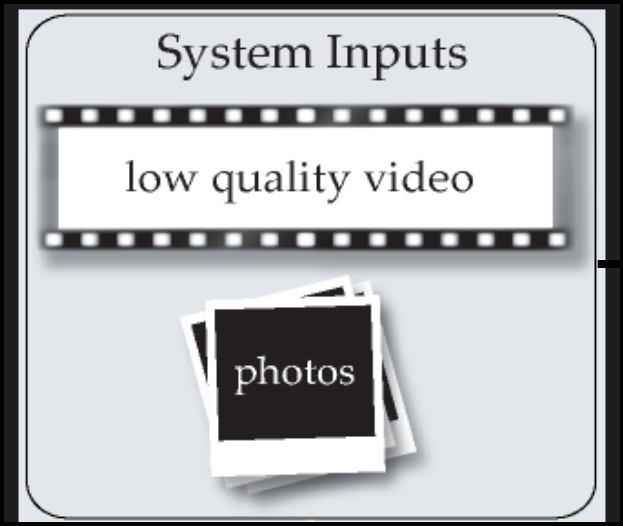
- One framework. Variety of enhancements:
  - Transferring photographic qualities
    - resolution
    - exposure
    - dynamic range
  - Easy video authoring
    - object touch-up
    - image filters
    - object removal
    - camera shake removal

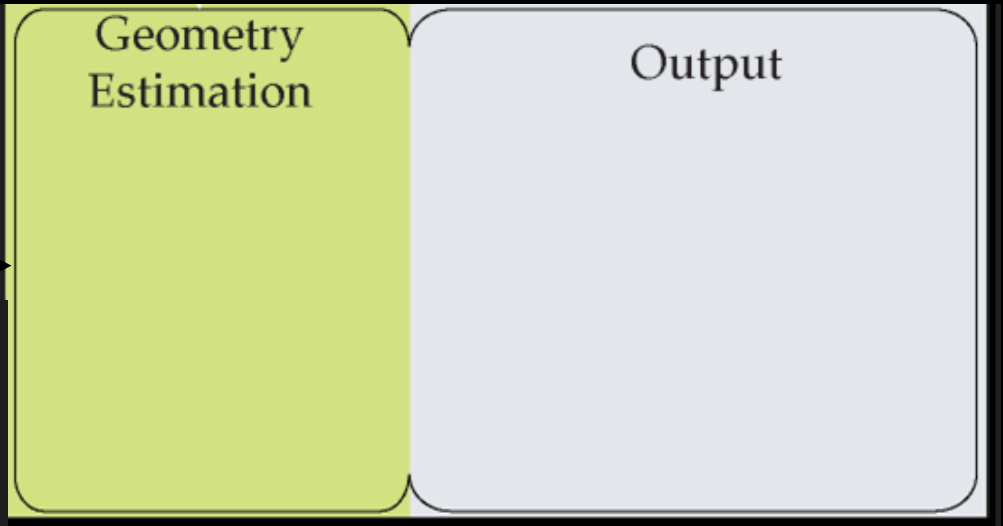
System Overview

System Overview

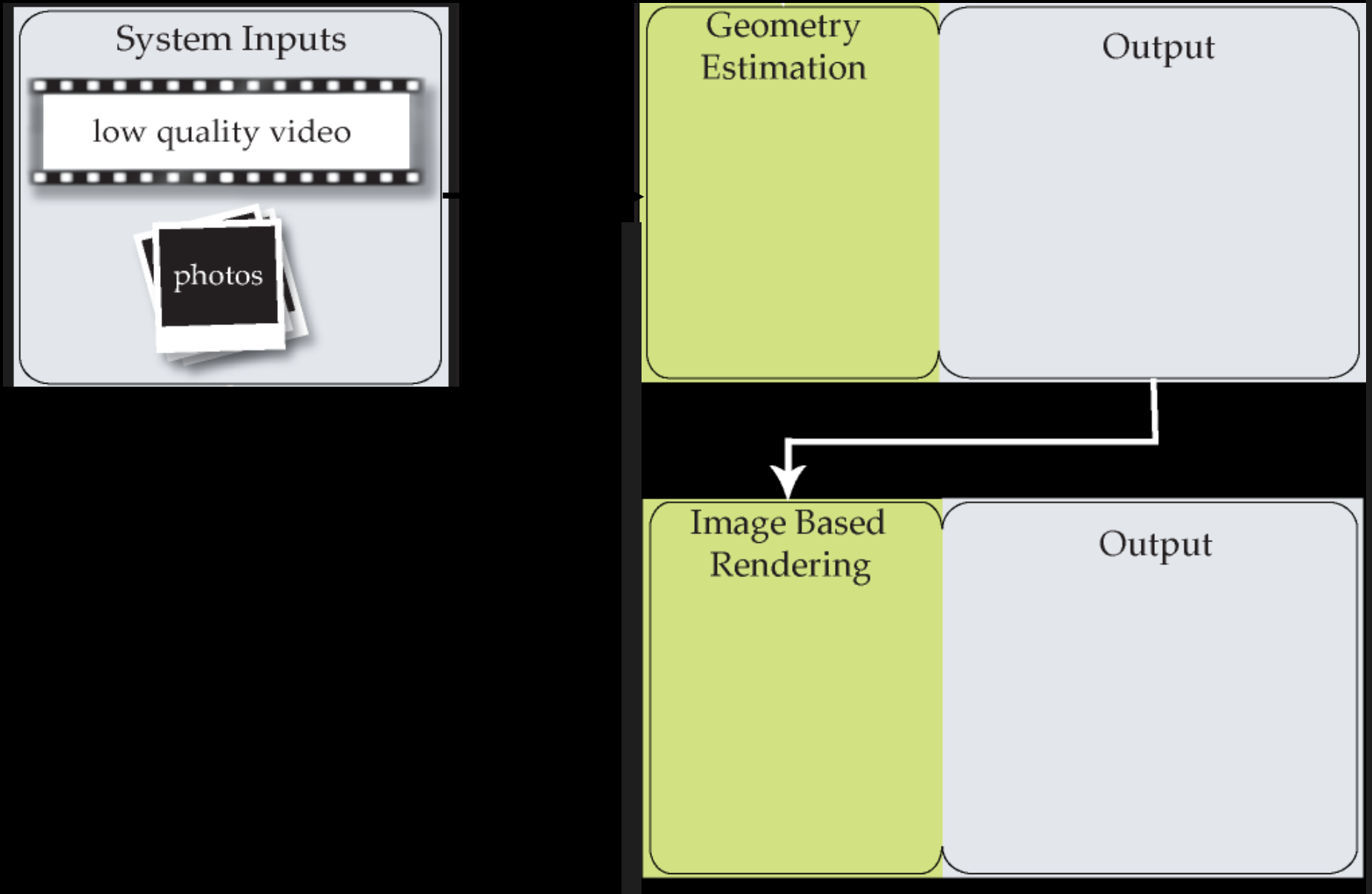


System Overview

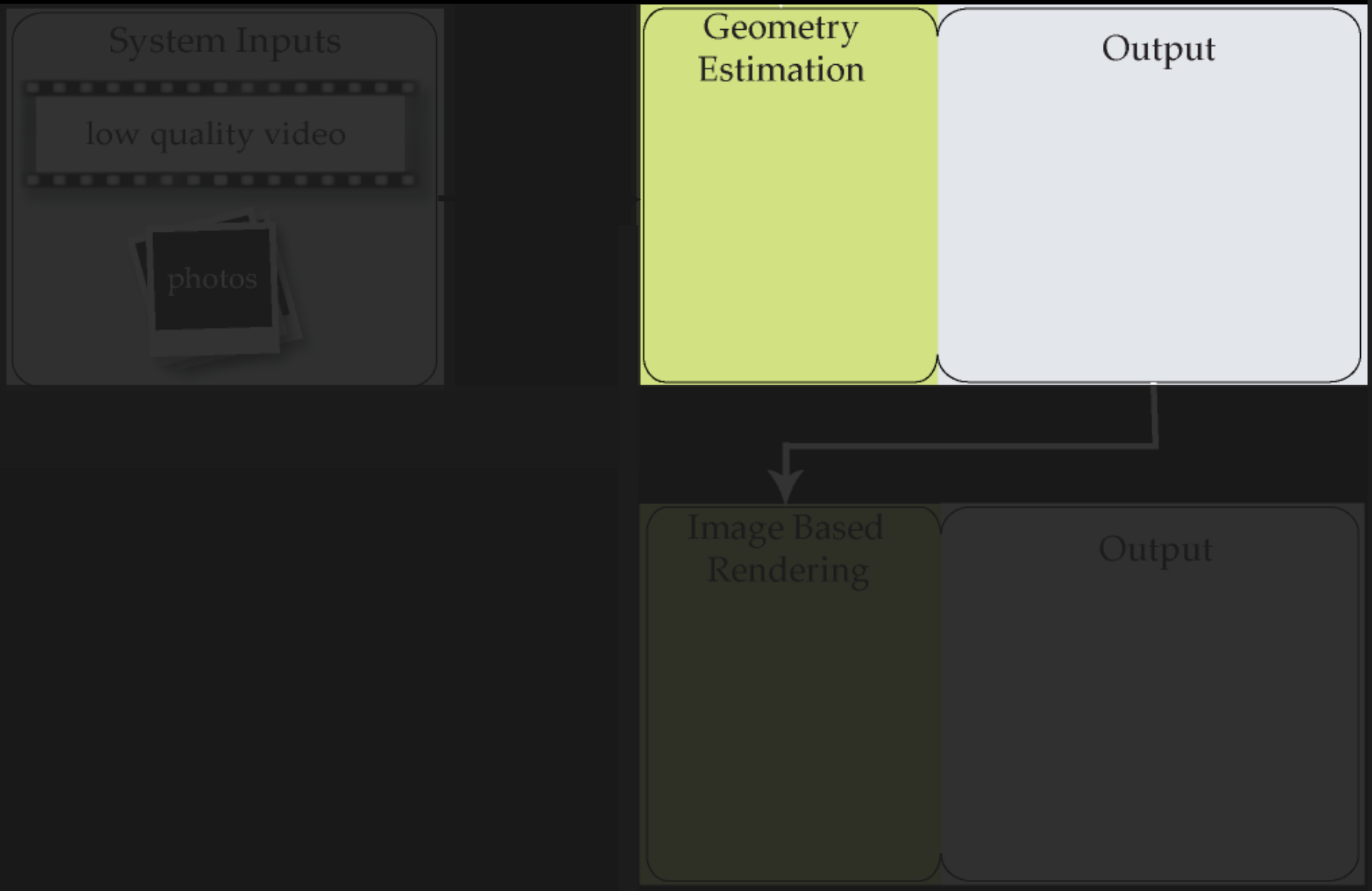




System Overview



System Overview



Geometry Estimation

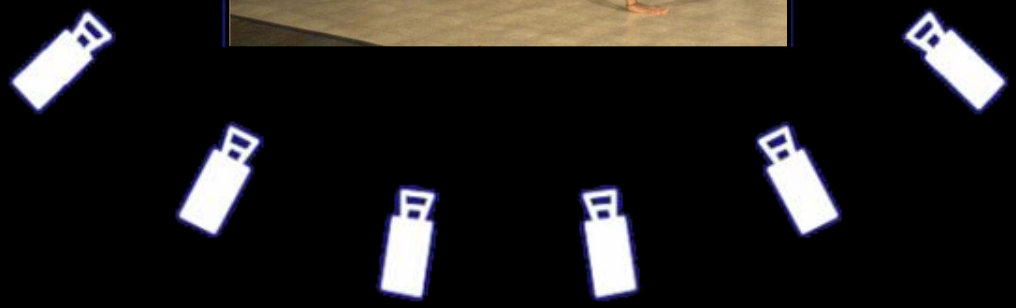
Structure from motion

Multi-View Stereo

Multi-View Stereo

- Based on Zitnick et al's work at SIGGRAPH '04
  - Good for videos and noisy images

Multi-View Stereo



synchronized camera arrays

Multi-View Stereo

- Based on Zitnick et al's work at SIGGRAPH '04

over-segmentation based approach

Multi-View Stereo

- Based on Zitnick et al's work at SIGGRAPH '04

over-segmentation based approach

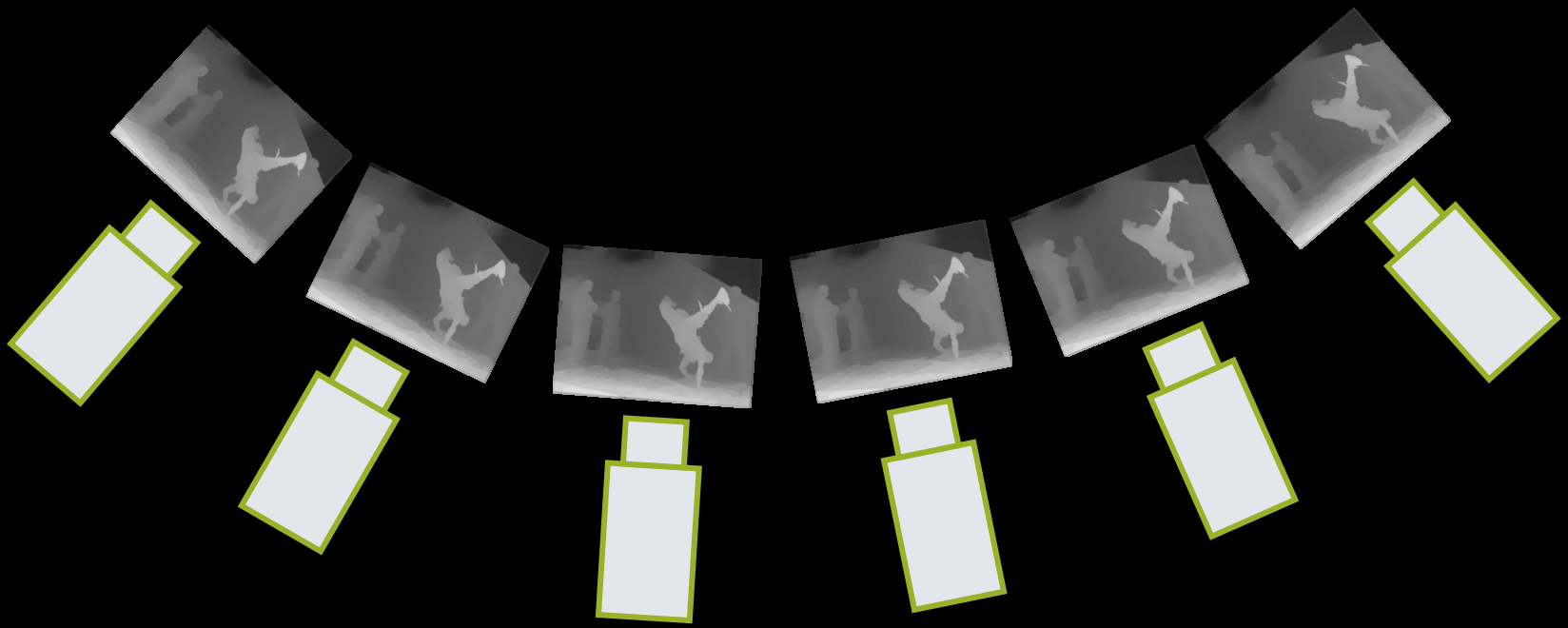
Multi-View Stereo

- Based on Zitnick et al's work at SIGGRAPH '04

BP based optimization

Multi-View Stereo

- Based on Zitnick et al's work at SIGGRAPH '04



depth consistency across views

Multi-View Stereo

- Our Contributions

Multi-View Stereo

- Our Contributions
  - Augmented depth planes
    - View-dependent planes
    - Non-fronto parallel planes

Multi-View Stereo

- Our Contributions
  - Augmented depth planes
    - View-dependent planes
    - Non-fronto parallel planes
  - Automated viewpoint graph construction

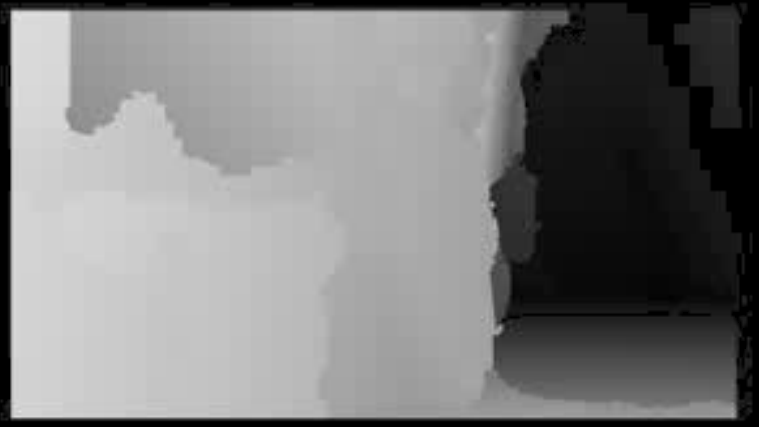
Multi-View Stereo

- Our Contributions
  - Augmented depth planes
    - View-dependent planes
    - Non-fronto parallel planes
  - Automated viewpoint graph construction
  - Using SFM point cloud as a soft prior

Geometry Estimation

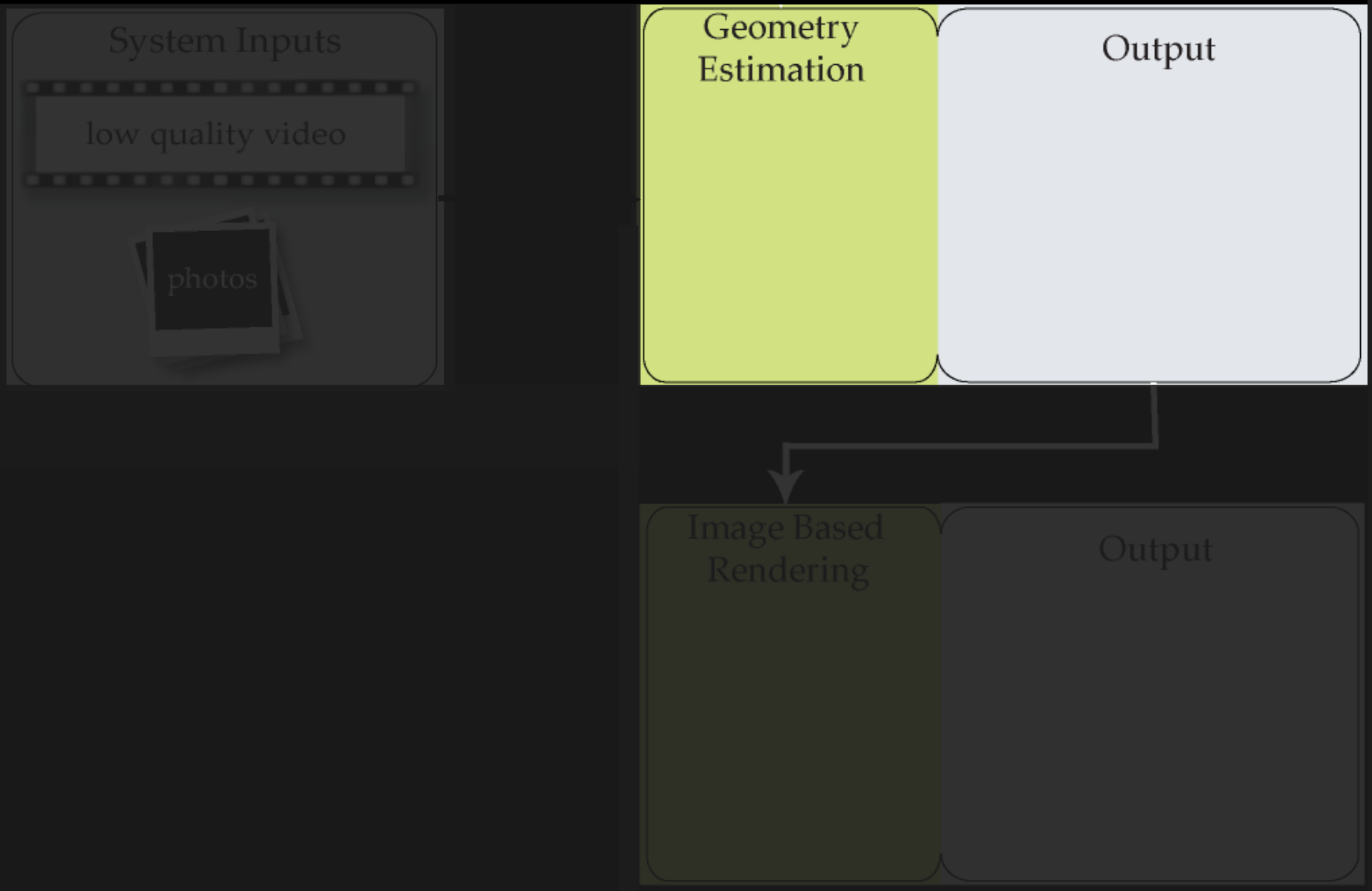
Multi-view stereo

Input video

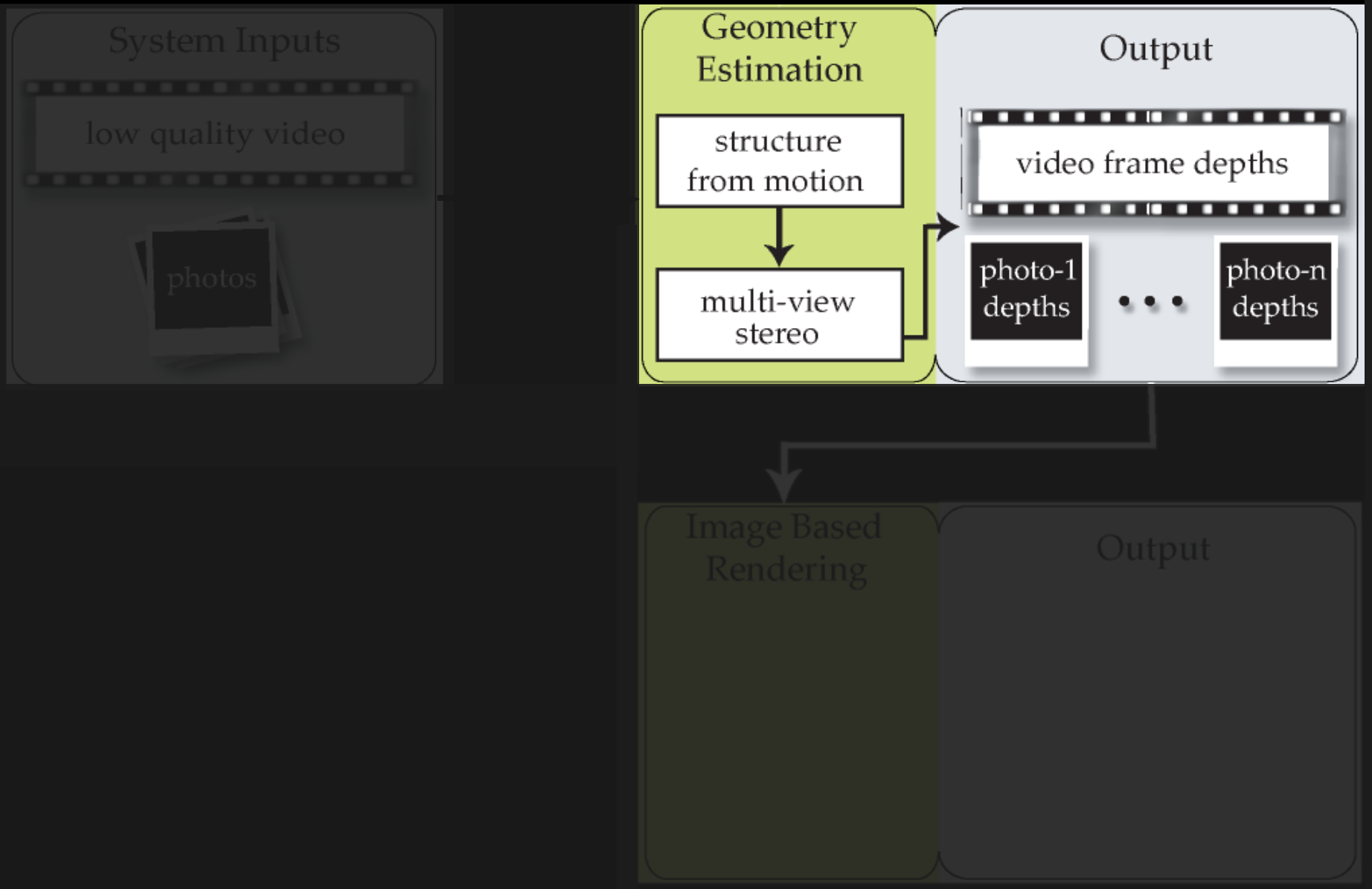


Depth maps

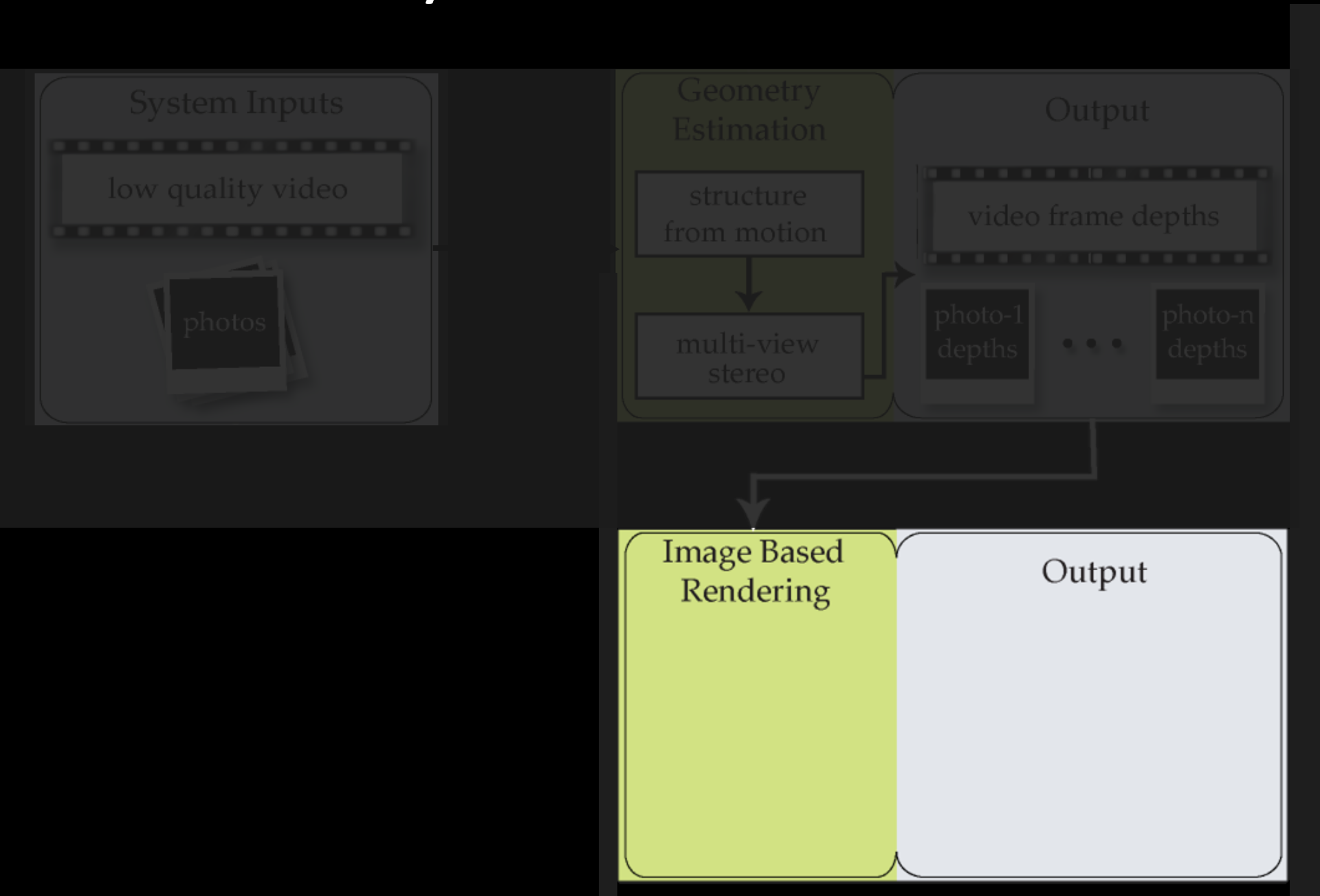
System Overview



System Overview

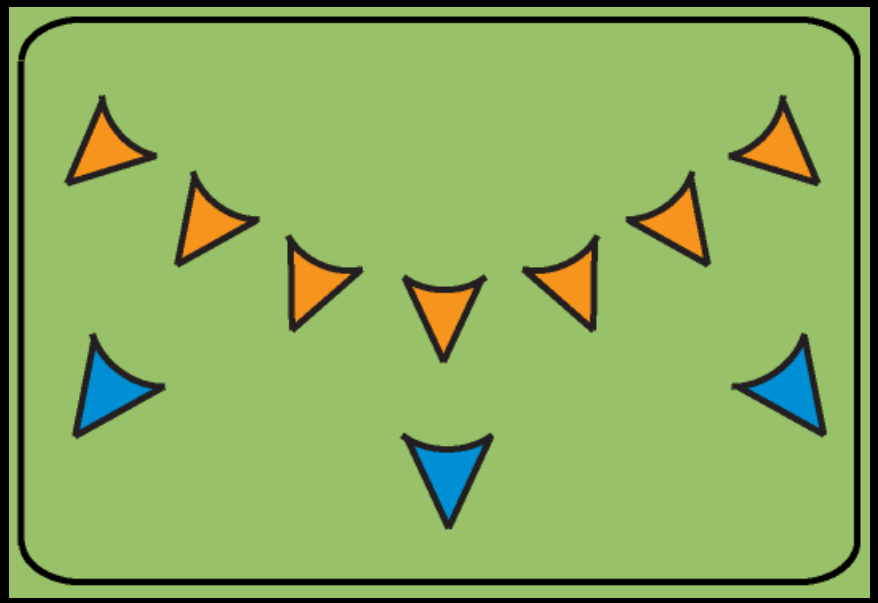


System Overview



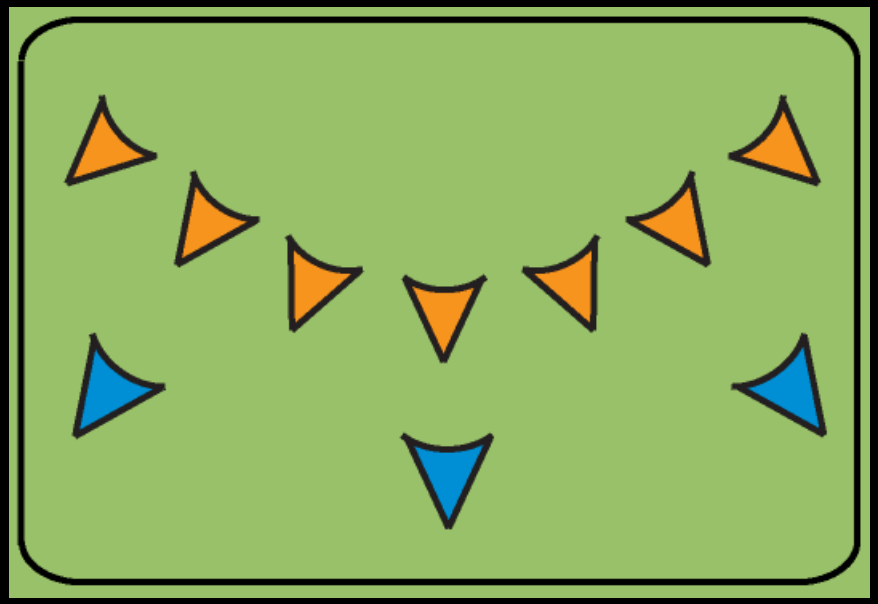
Novel View Interpolation

Novel View Interpolation



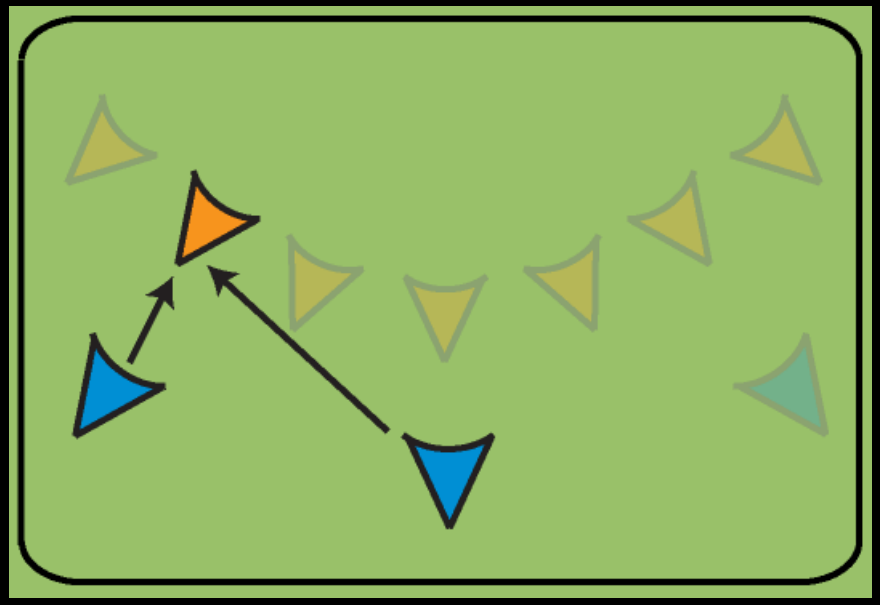
Novel View Interpolation

- Objective
  - Reconstruct every video frame (orange cones) using nearby photos (blue cones)



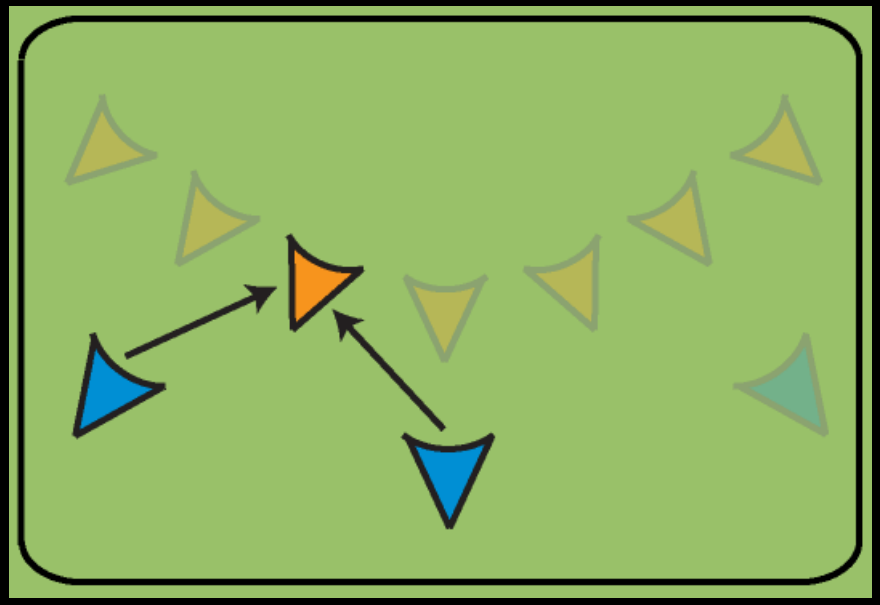
Novel View Interpolation

- Objective
  - Reconstruct every video frame (orange cones) using nearby photos (blue cones)



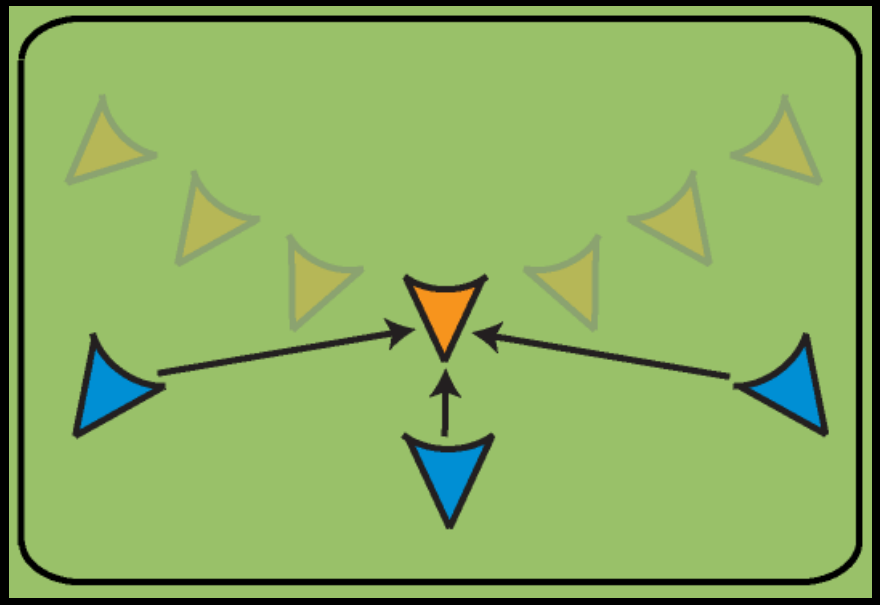
Novel View Interpolation

- Objective
  - Reconstruct every video frame (orange cones) using nearby photos (blue cones)



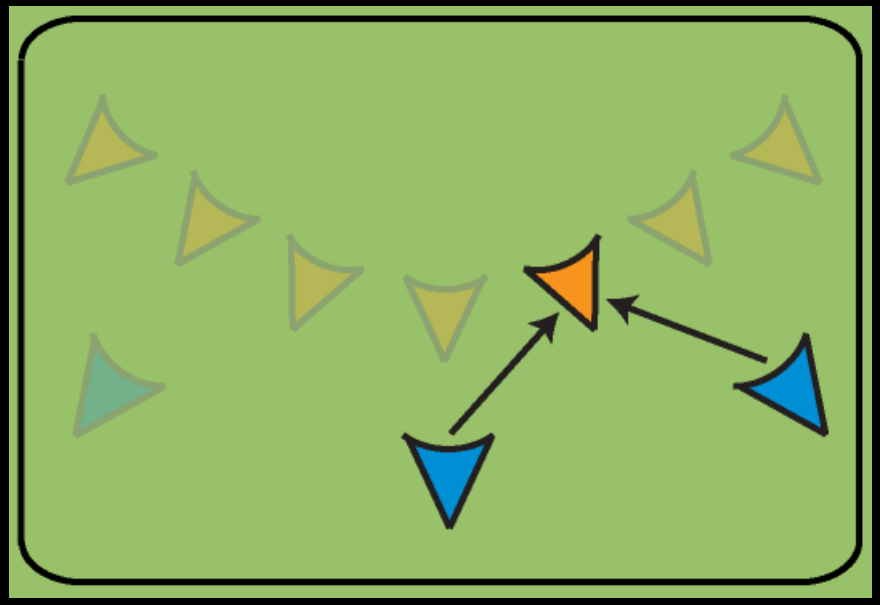
Novel View Interpolation

- Objective
  - Reconstruct every video frame (orange cones) using nearby photos (blue cones)



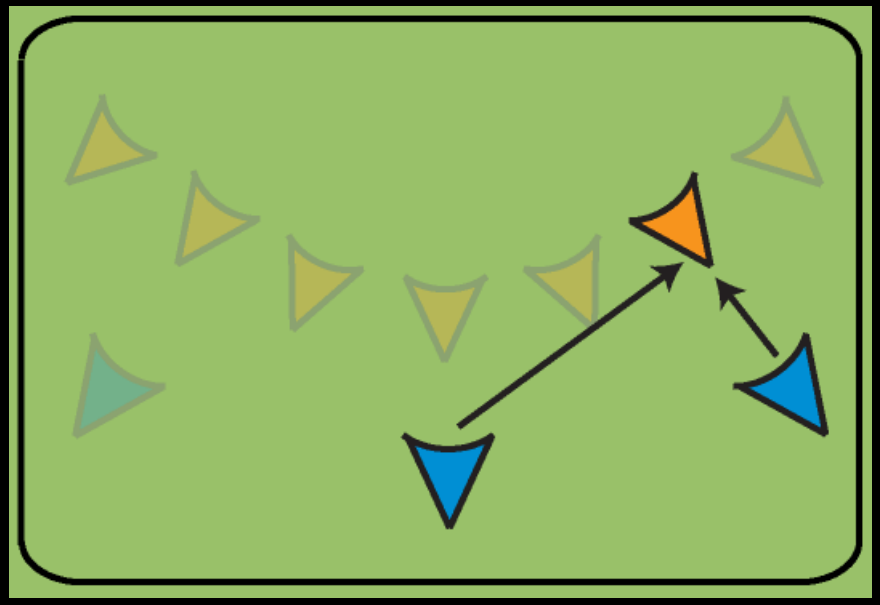
Novel View Interpolation

- Objective
  - Reconstruct every video frame (orange cones) using nearby photos (blue cones)



Novel View Interpolation

- Objective
  - Reconstruct every video frame (orange cones) using nearby photos (blue cones)



Novel View Interpolation

Video Frame

Novel View Interpolation

Video Frame

Photo-1

Photo-2

Novel View Interpolation

Video Frame

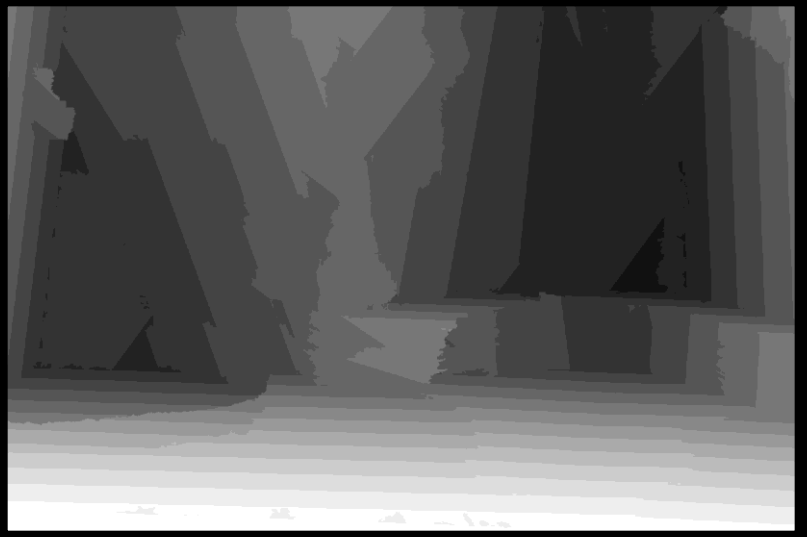


Photo-1 Depths

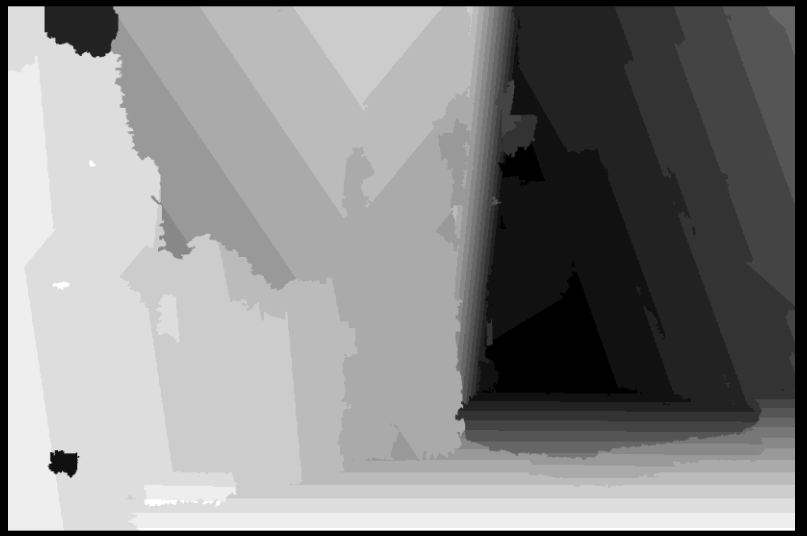


Photo-2 Depths

Novel View Interpolation

Projection-1

Video Frame

Projection-2

Novel View Interpolation

- Simple IBR Reconstruction
  - Result = average(Projection-1, Projection-2)

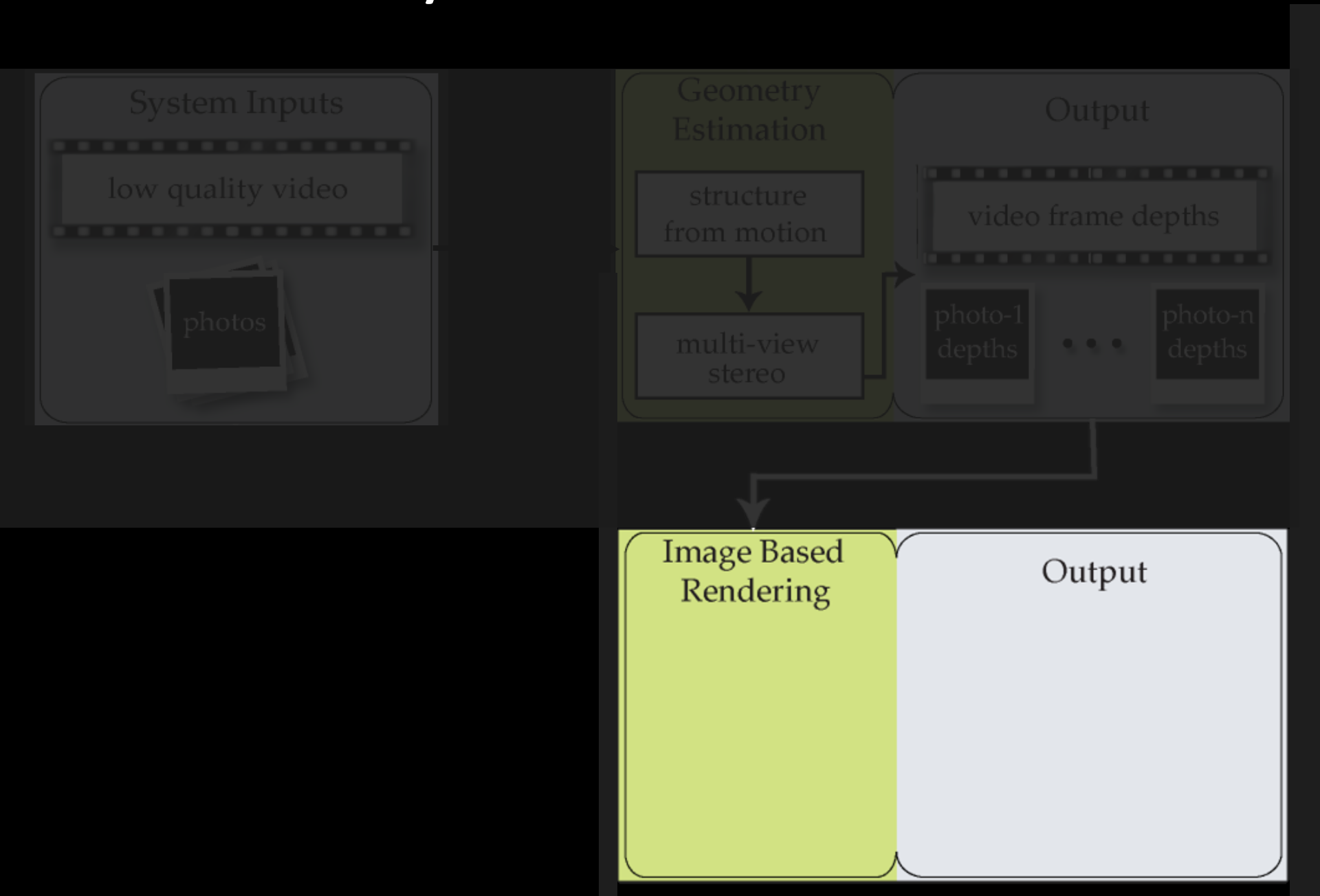
Novel View Interpolation

- Simple IBR Reconstruction
  - Result = average(Projection-1, Projection-2)
  - Artifacts
    - blurring and ghosting
    - loss of dynamic lighting

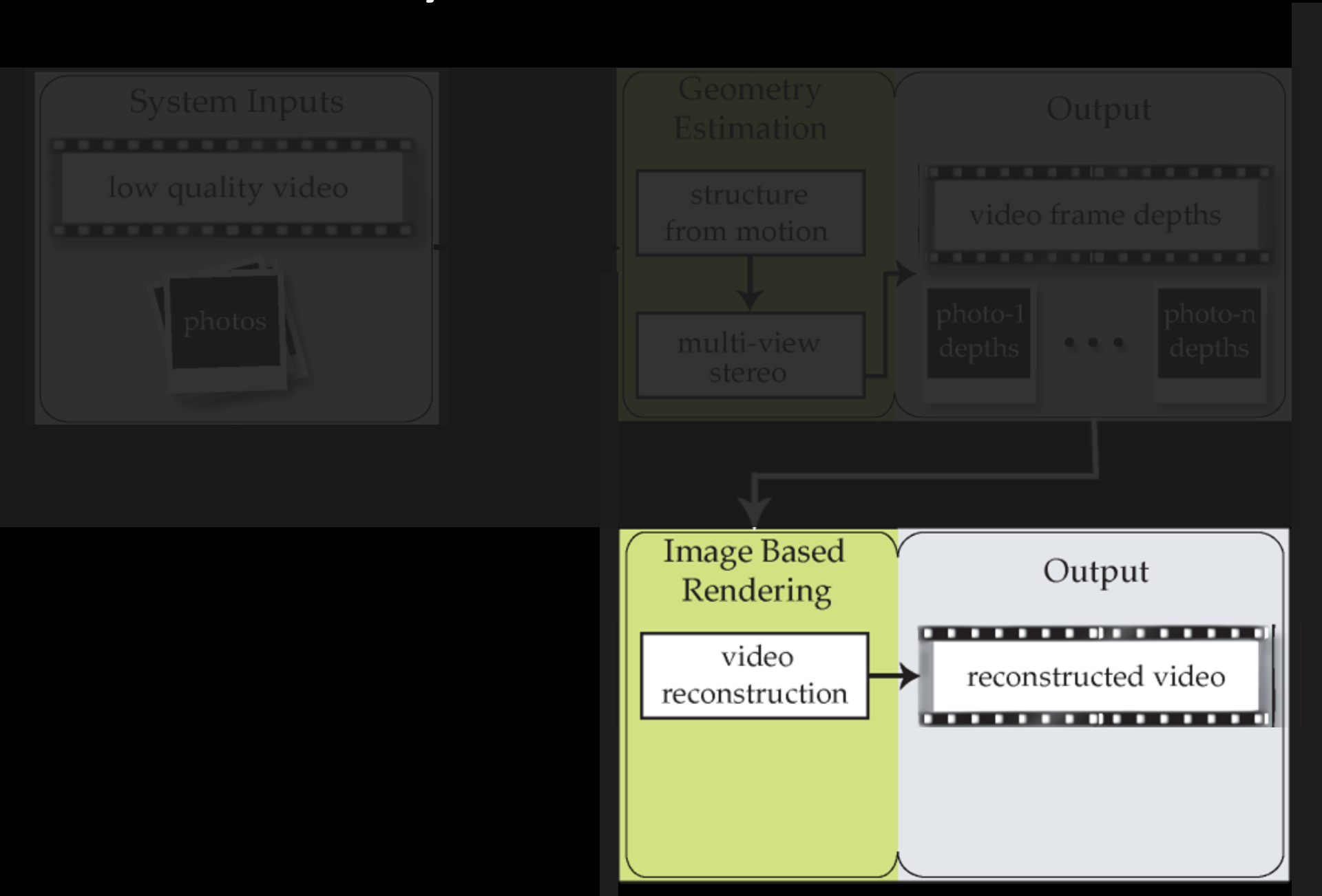
Novel View Interpolation

- Simple IBR Reconstruction
  - Result = average(Projection-1 , Projection-2)
  - Artifacts
    - blurring and ghosting
    - loss of dynamic lighting
- Our Solution
  - Patch based reconstruction
  - Gradient domain compositing
  - Exploit video data

System Overview



System Overview



Video Reconstruction

- Objective
  - Reconstruct every video frame using projected images

Video Reconstruction

- Our Solution
  - Reconstruction problem = Labeling problem
  - Each projection provides one label
  - Graphcuts optimization to find a labeling that minimizes a well-designed cost function

Video Reconstruction

Video Frame

Projection-1

Projection-2

Video Reconstruction

Label-1

Video Frame

Label-2

Video Reconstruction

Labeling

Label-1

Label-2

Video Reconstruction

- Cost function:

Video Reconstruction

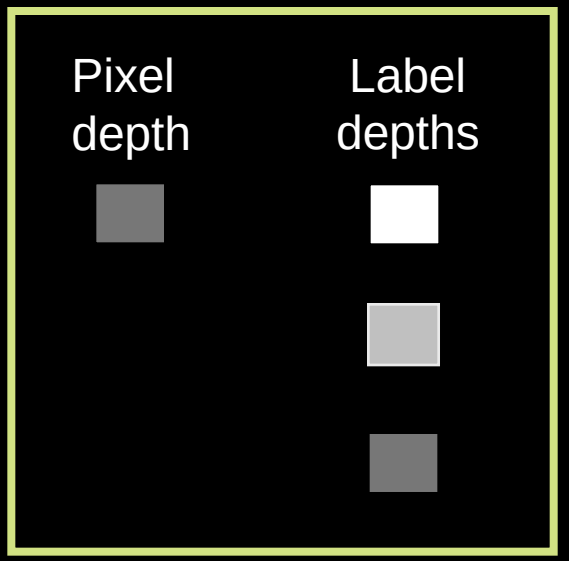
- Cost function:
  - Data term encourages:

Video Reconstruction

- Cost function:
  - Data term encourages:
    - depth matching

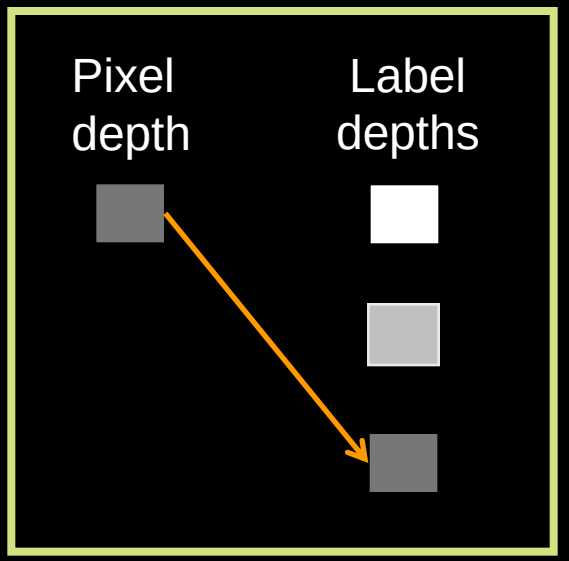
Video Reconstruction

- Cost function:
  - Data term encourages:
    - depth matching



Video Reconstruction

- Cost function:
  - Data term encourages:
    - depth matching

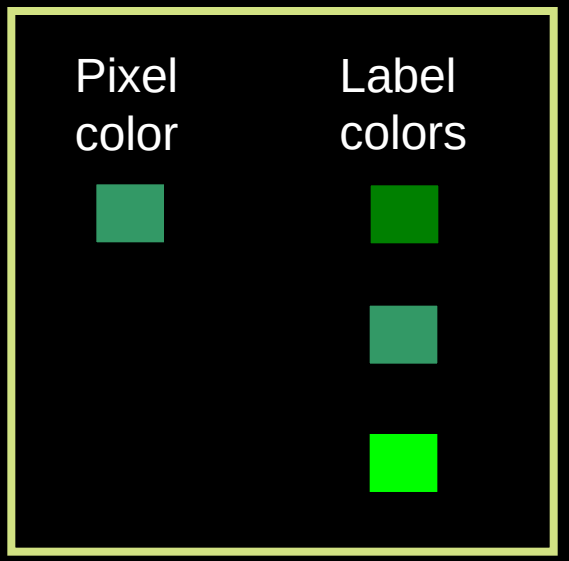


Video Reconstruction

- Cost function:
  - Data term encourages:
    - depth matching
    - color matching

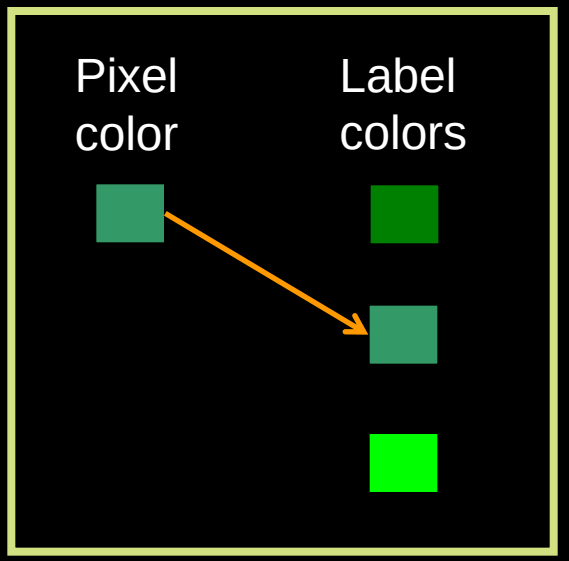
Video Reconstruction

- Cost function:
  - Data term encourages:
    - depth matching
    - color matching



Video Reconstruction

- Cost function:
  - Data term encourages:
    - depth matching
    - color matching

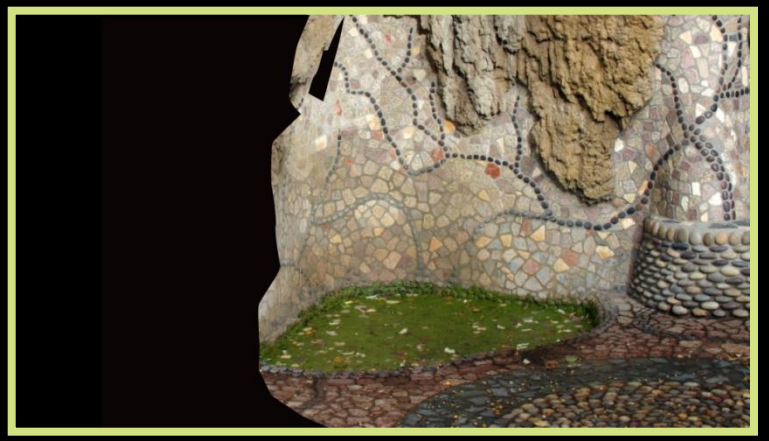


Video Reconstruction

- Cost function:
  - Data term encourages:
    - depth matching
    - color matching
    - no holes

Video Reconstruction

- Cost function:
  - Data term encourages:
    - depth matching
    - color matching
    - no holes



Video Reconstruction

- Cost function:
  - Data term encourages:
    - depth matching
    - color matching
    - no holes

Video Reconstruction

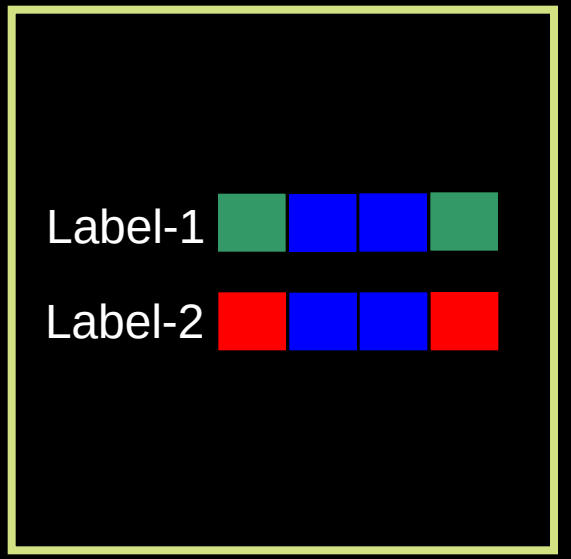
- Cost function:
  - Data term encourages:
    - depth matching
    - color matching
    - no holes
  - Smoothness term encourages:
    - Seamless label transitions
    - Seams to run:

Video Reconstruction

- Cost function:
  - Data term encourages:
    - depth matching
    - color matching
    - no holes
  - Smoothness term encourages:
    - Seamless label transitions
    - Seams to run:
      - through regions of similar color

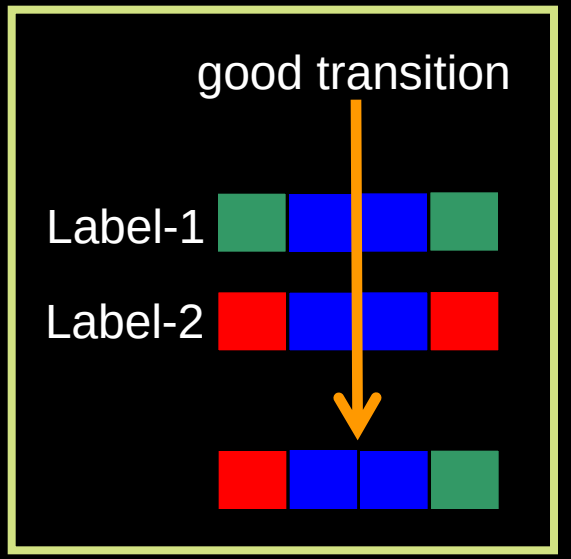
Video Reconstruction

- Cost function:
  - Data term encourages:
    - depth matching
    - color matching
    - no holes
  - Smoothness term encourages:
    - Seamless label transitions
    - Seams to run:
      - through regions of similar color



Video Reconstruction

- Cost function:
  - Data term encourages:
    - depth matching
    - color matching
    - no holes
  - Smoothness term encourages:
    - Seamless label transitions
    - Seams to run:
      - through regions of similar color

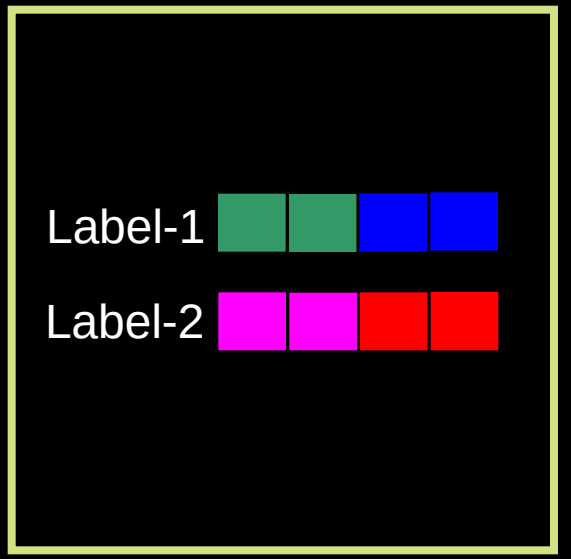


Video Reconstruction

- Cost function:
  - Data term encourages:
    - depth matching
    - color matching
    - no holes
  - Smoothness term encourages:
    - Seamless label transitions
    - Seams to run:
      - through regions of similar color
      - along strong edges

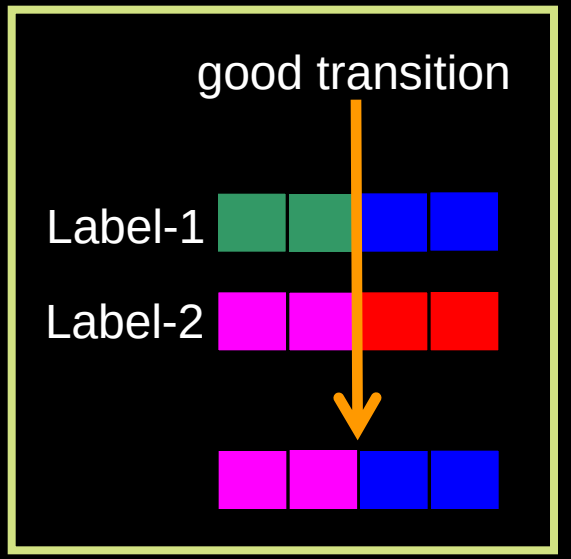
Video Reconstruction

- Cost function:
  - Data term encourages:
    - depth matching
    - color matching
    - no holes
  - Smoothness term encourages:
    - Seamless label transitions
    - Seams to run:
      - through regions of similar color
      - along strong edges



Video Reconstruction

- Cost function:
  - Data term encourages:
    - depth matching
    - color matching
    - no holes
  - Smoothness term encourages:
    - Seamless label transitions
    - Seams to run:
      - through regions of similar color
      - along strong edges



Video Reconstruction

Labeling

Label-1

Label-2

Video Reconstruction

Label-1

Label-2

Video Reconstruction

Label-1

Reconstructed
Video Frame

Label-2

Video Reconstruction

Label-1

Video Frame

Label-2

Video Reconstruction

Label-1

Reconstructed
Video Frame

Label-2

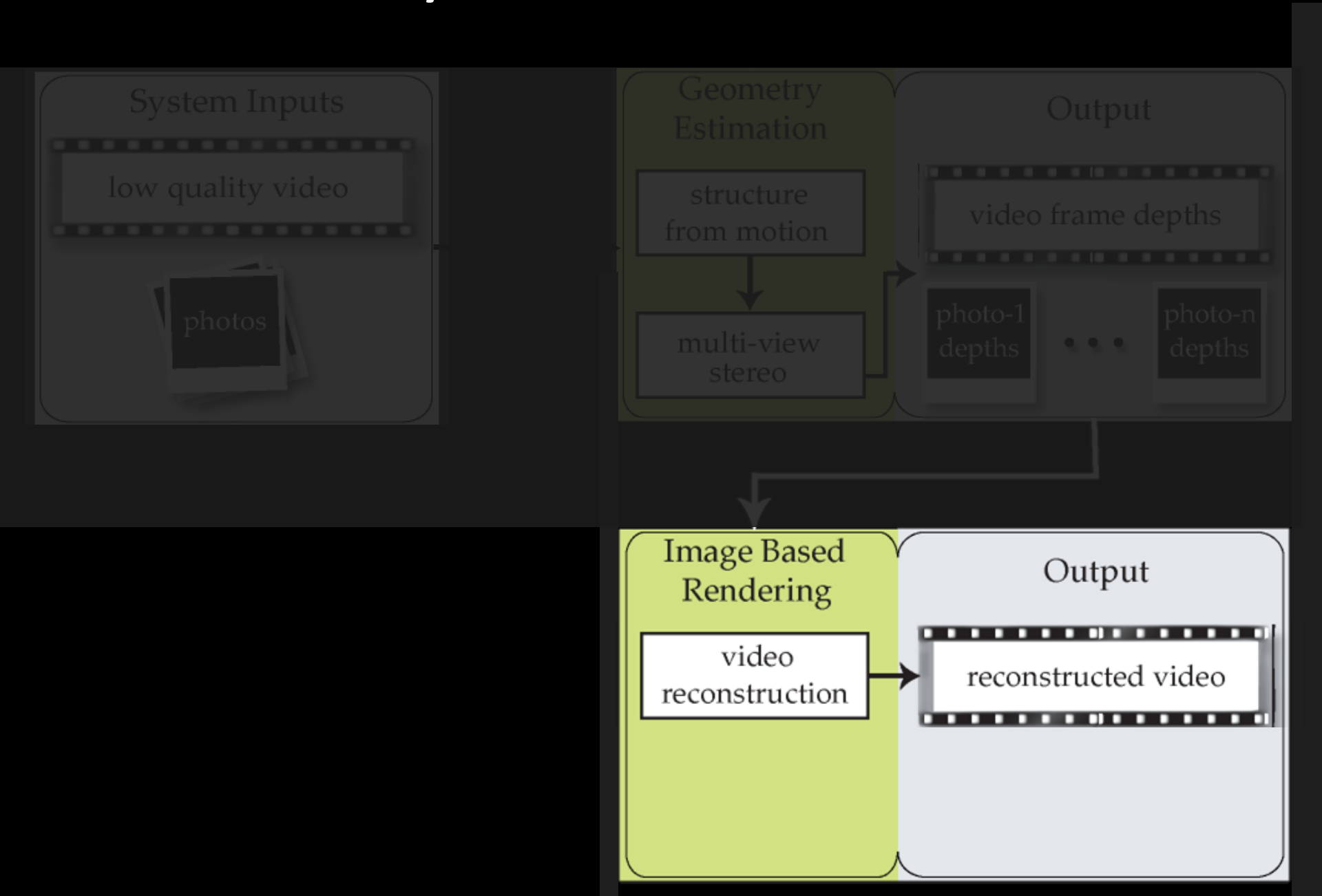
Video reconstruction

Original video

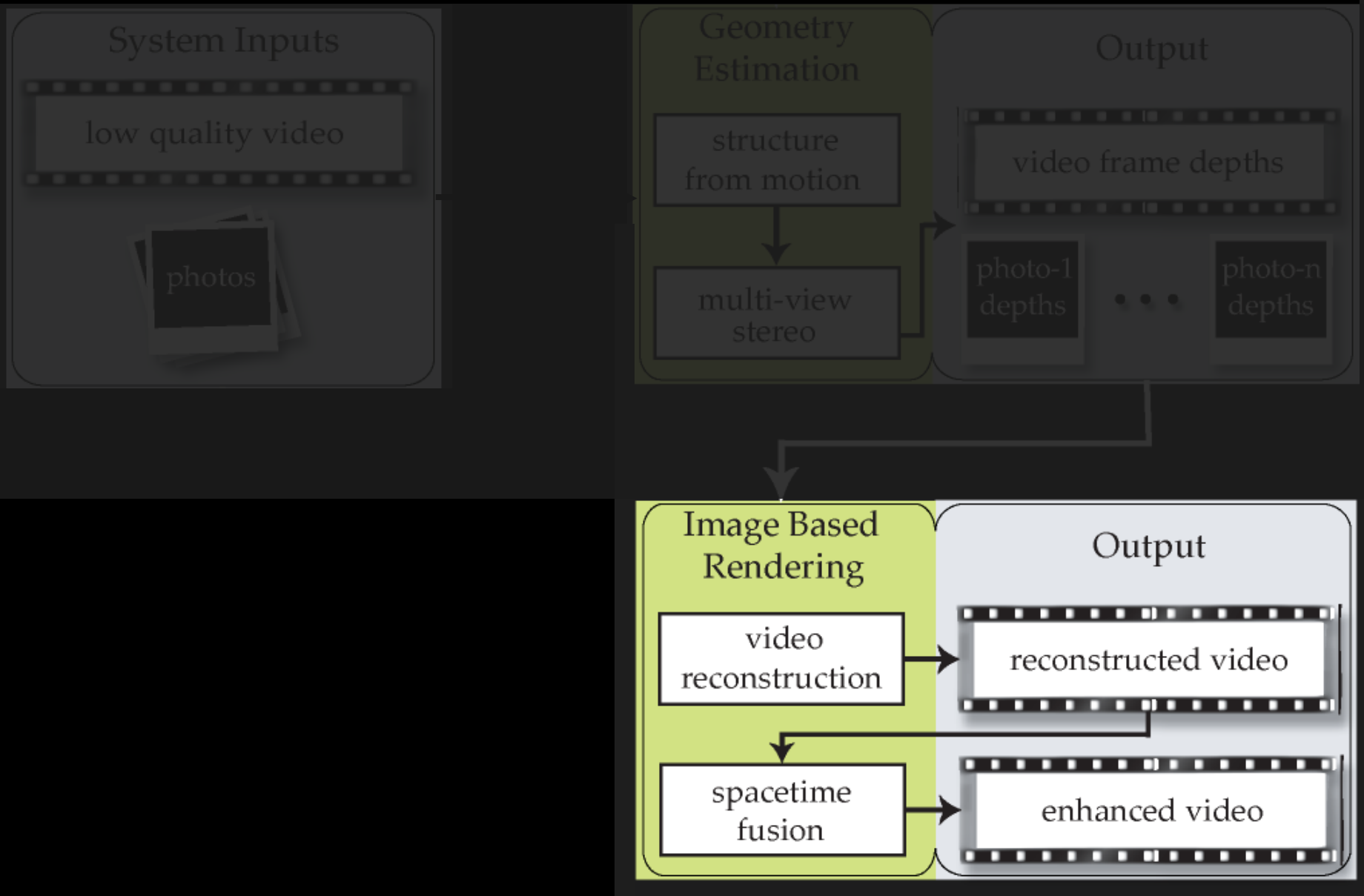
Video reconstruction

Reconstructed video

System Overview



System Overview



Spacetime Fusion

- Objective: To eliminate
  - Holes
  - Spatial seams
  - Temporal incoherence
  - Loss of lighting dynamics

Spacetime Fusion

- Objective: To eliminate
  - Holes
  - Spatial seams
  - Temporal incoherence
  - Loss of lighting dynamics
- Solution
  - Define artifact free gradient field
  - Integrate gradient field

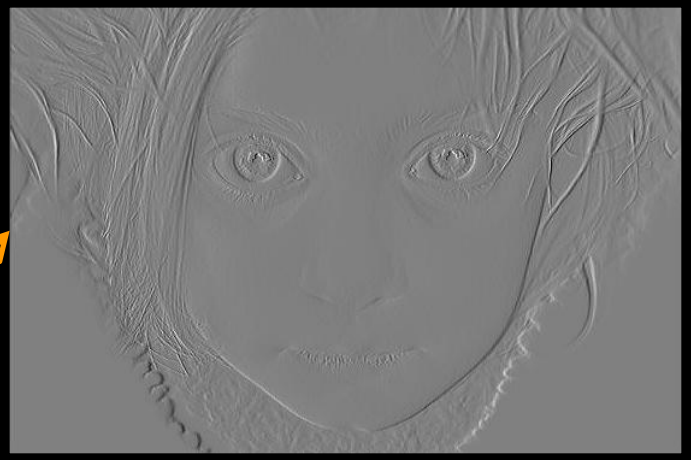
Gradient Field Integration

Gradient Field Integration

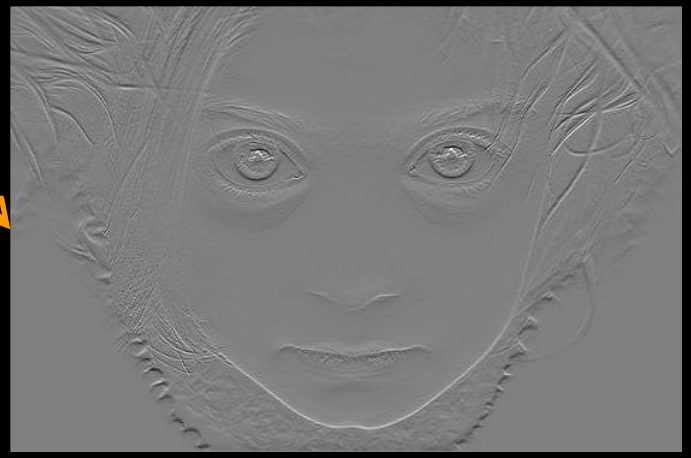
I

Gradient Field Integration

I



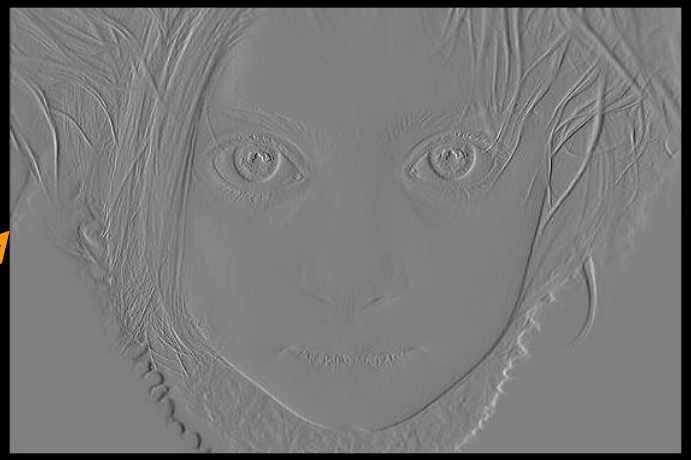
G_x



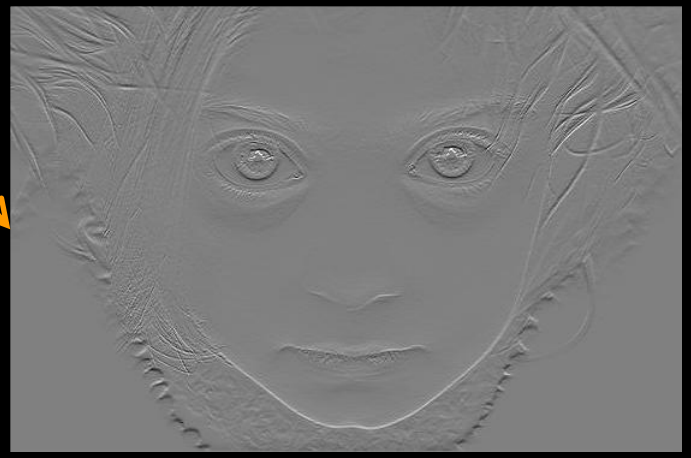
G_y

Gradient Field Integration

I



G_x

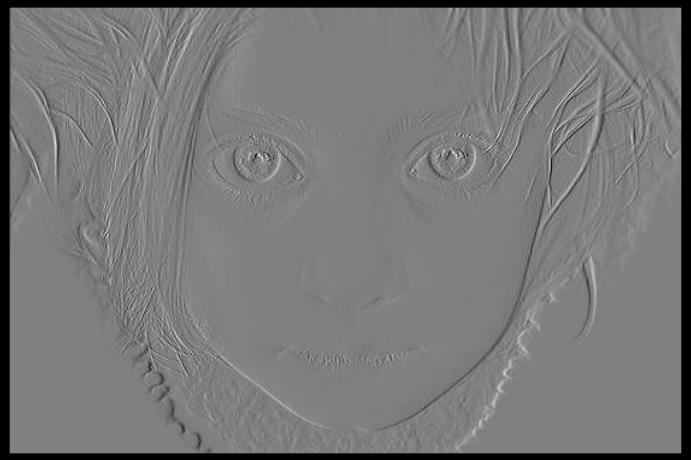


G_y

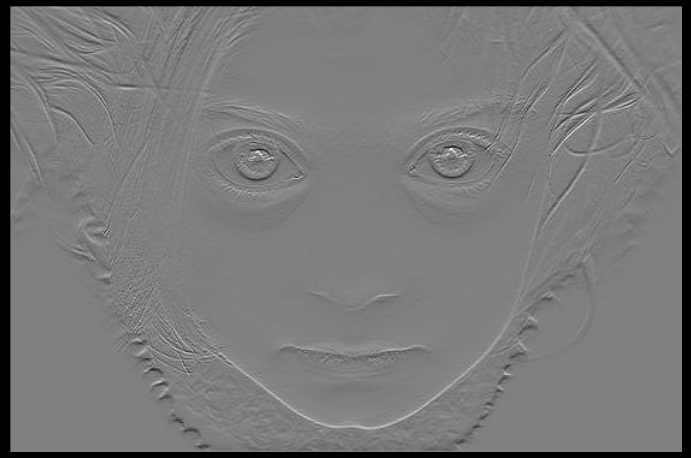
$$G_x(x, y) = I(x, y) - I(x - 1, y)$$

$$G_y(x, y) = I(x, y) - I(x, y - 1)$$

Gradient Field Integration



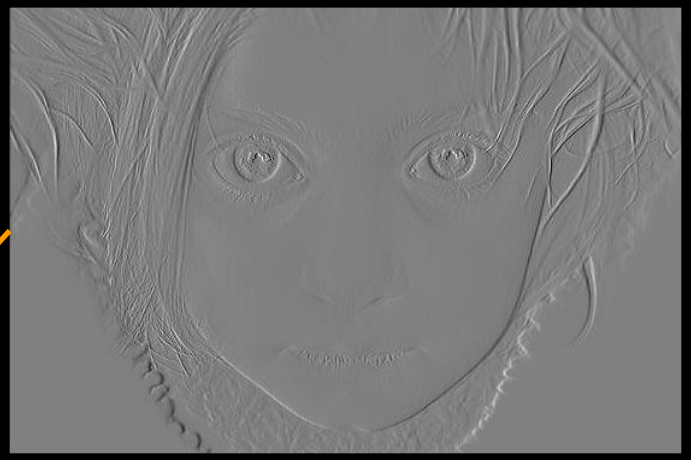
G_x



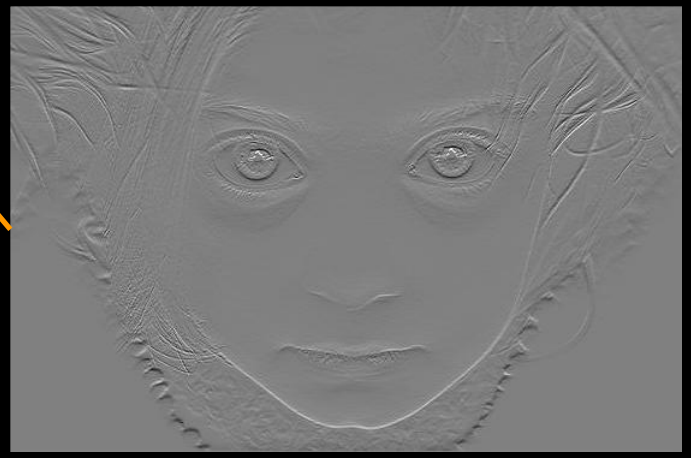
G_y

Gradient Field Integration

I

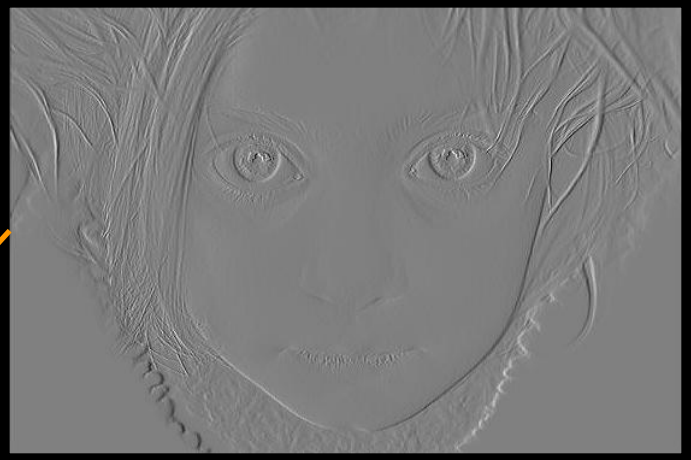


G_x



G_y

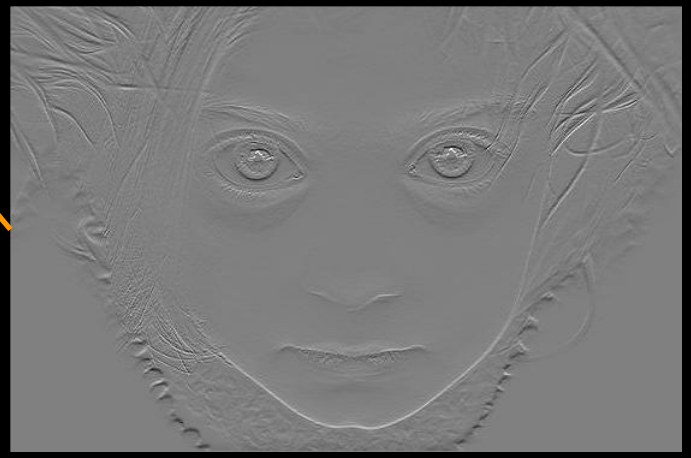
Gradient Field Integration



G_x

?

I

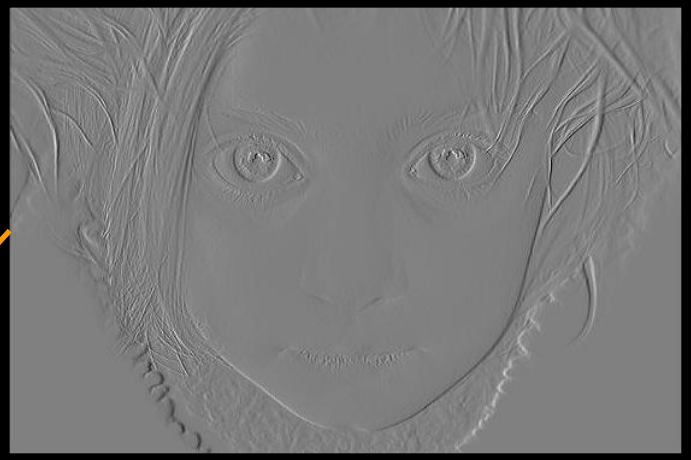


G_y

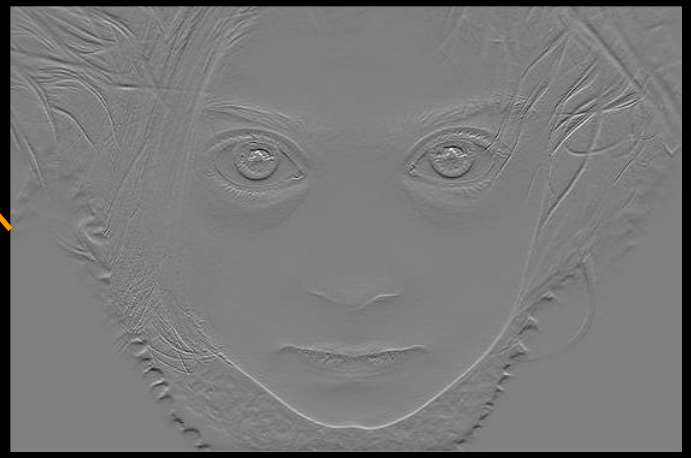
Solve linear system:

$\boldsymbol{Av} = \boldsymbol{b}$

Gradient Field Integration



G_x



G_y

?

I

Solve linear system:

$\boldsymbol{Av} = \boldsymbol{b}$

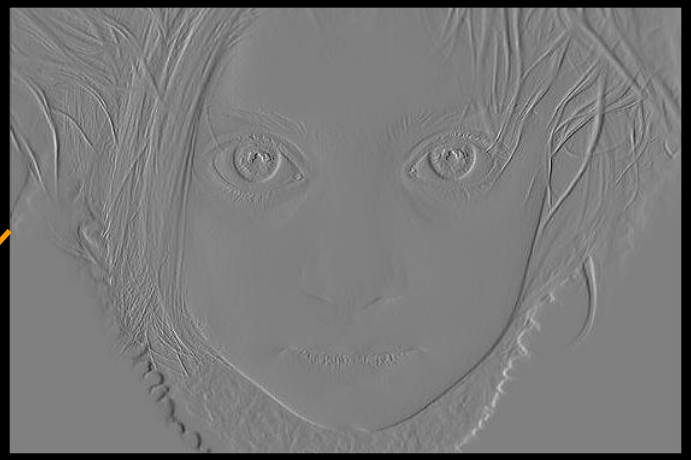
Constraints:

$\boldsymbol{v}_{x,y} - \boldsymbol{v}_{x-1,y} = \boldsymbol{G}_x(x, y)$

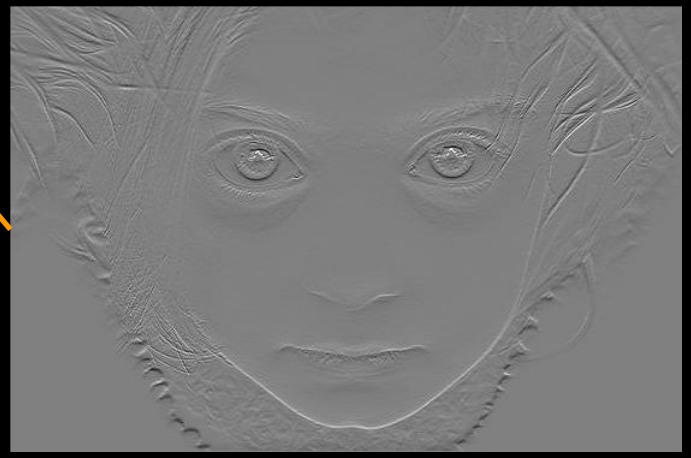
$\boldsymbol{v}_{x,y} - \boldsymbol{v}_{x,y-1} = \boldsymbol{G}_y(x, y)$

Gradient Field Integration

I



G_x



G_y

Solve linear system:

$Av = b$

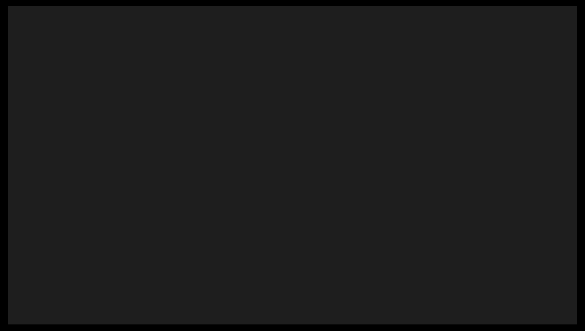
Constraints:

$v_{x,y} - v_{x-1,y} = G_x(x, y)$

$v_{x,y} - v_{x,y-1} = G_y(x, y)$

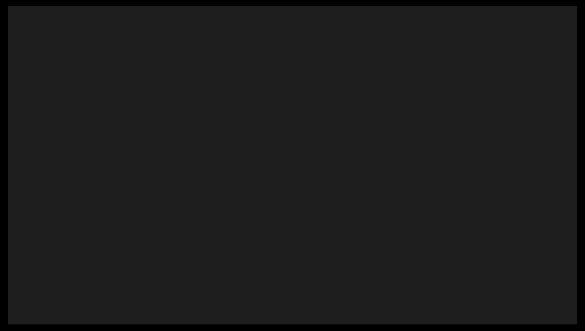
Defining Spatial Gradients

$G_{x,y}$



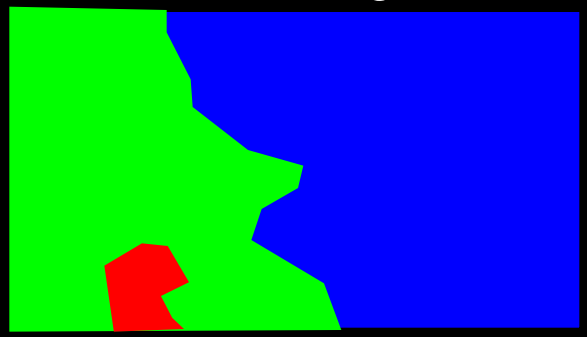
Defining Spatial Gradients

$G_{x,y}$

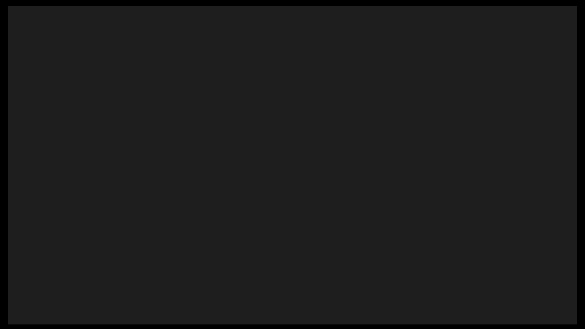


Defining Spatial Gradients

Labeling

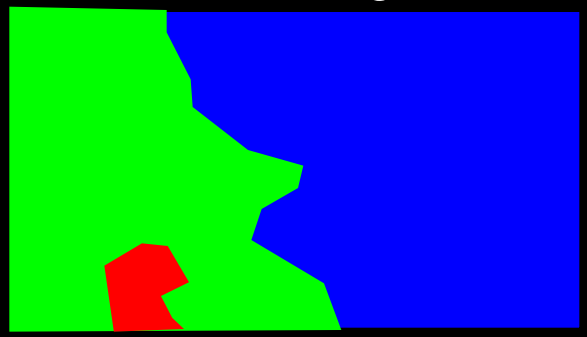


$G_{x,y}$



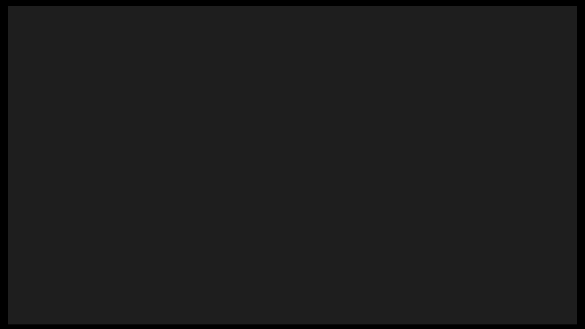
Defining Spatial Gradients

Labeling



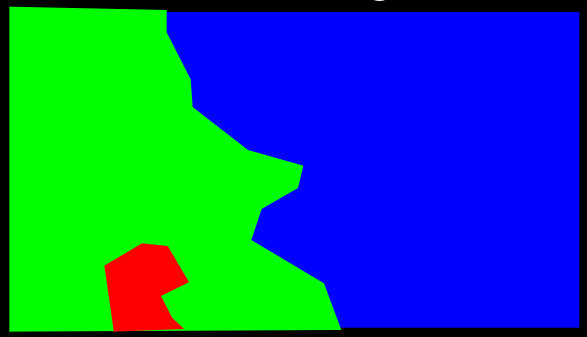
Label-1

$G_{x,y}$

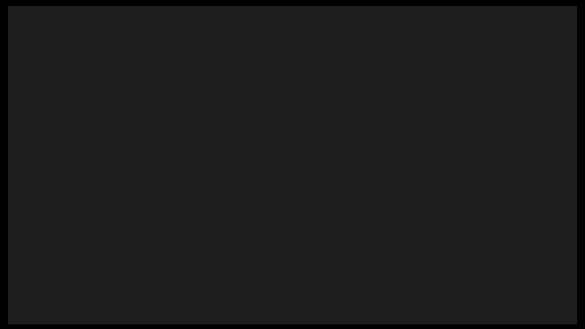


Defining Spatial Gradients

Labeling



$G_{x,y}$

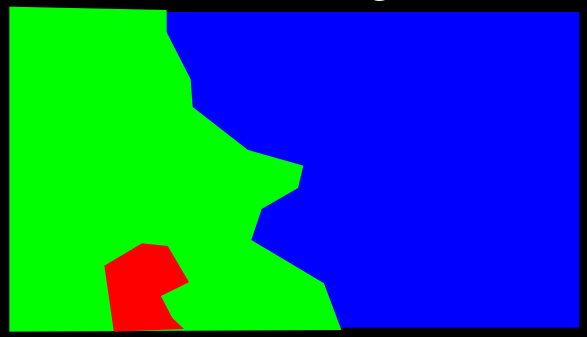


Label-1

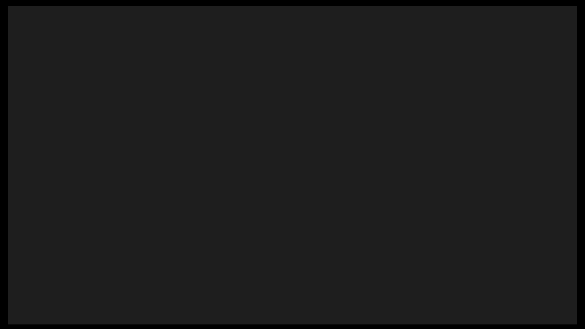
Label-2

Defining Spatial Gradients

Labeling



$G_{x,y}$



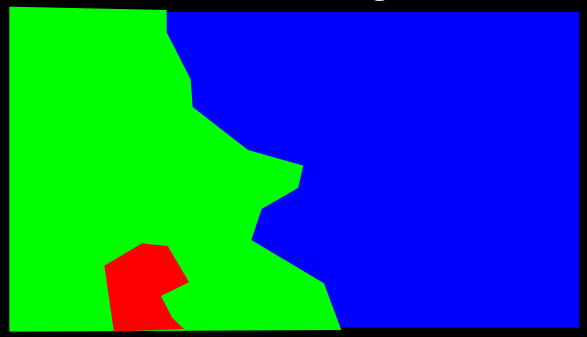
Label-1

Label-2

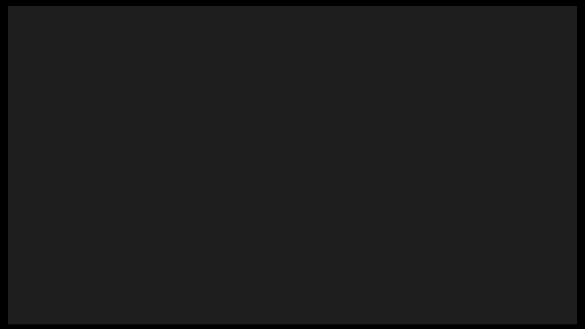
Video frame

Defining Spatial Gradients

Labeling



$G_{x,y}$



Label-1

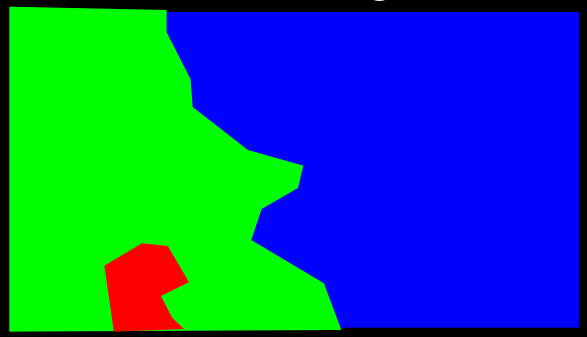
Label-2

Video frame

Inside a patch

Defining Spatial Gradients

Labeling



$G_{x,y}$

Label-1

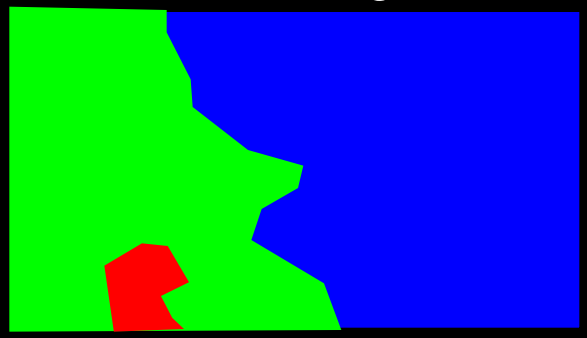
Label-2

Video frame

Inside a patch

Defining Spatial Gradients

Labeling



$G_{x,y}$

Label-1

Label-2

Video frame

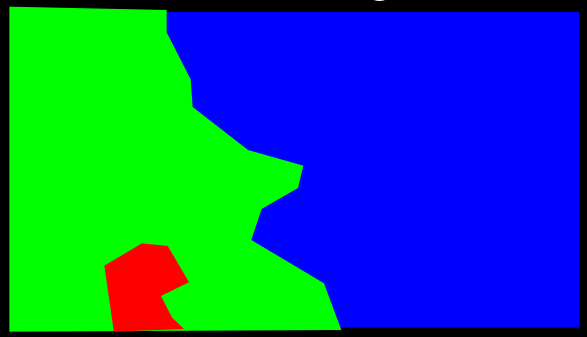
Inside a patch

Defining Spatial Gradients

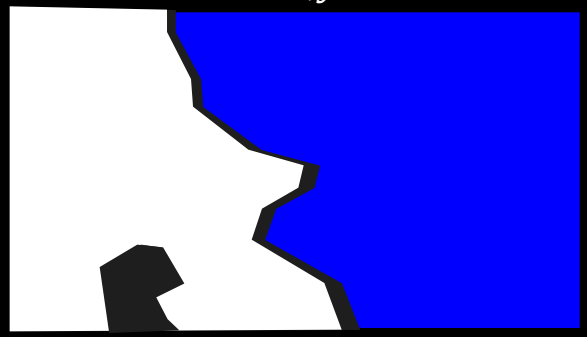
Inside a patch

Defining Spatial Gradients

Labeling



$G_{x,y}$



Label-1

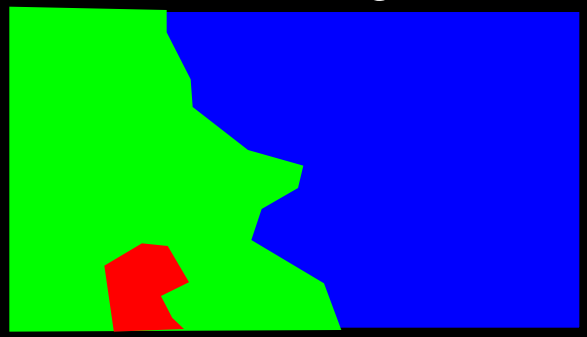
Label-2

Video frame

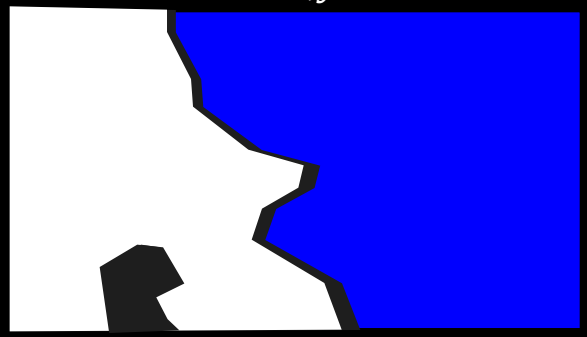
Inside a patch

Defining Spatial Gradients

Labeling



$G_{x,y}$



Label-1

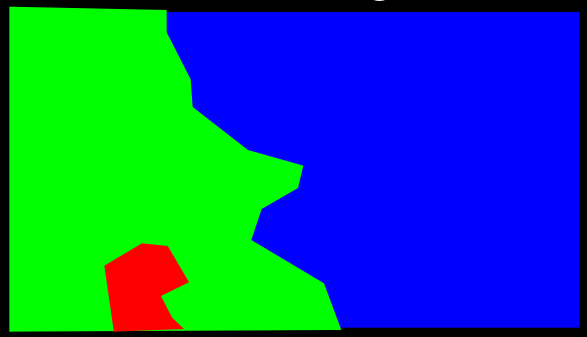
Label-2

Video frame

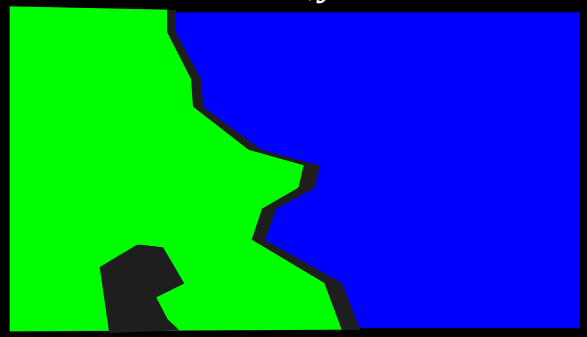
Inside a patch

Defining Spatial Gradients

Labeling



$G_{x,y}$



Label-1

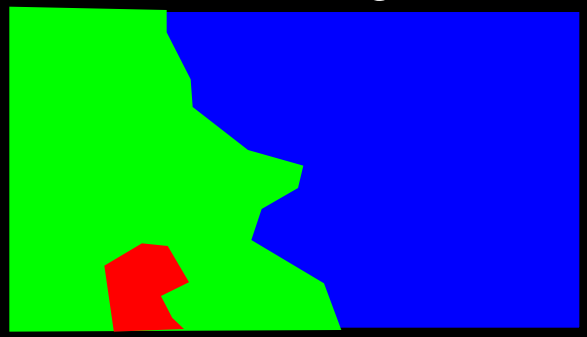
Label-2

Video frame

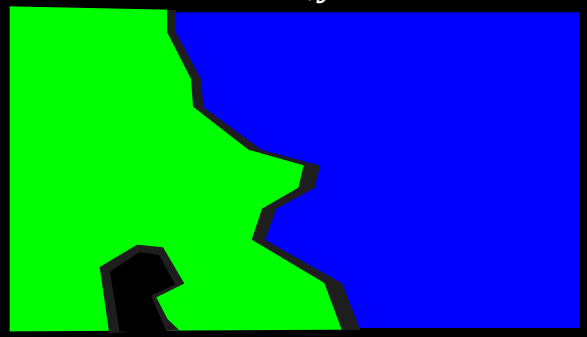
Inside a patch

Defining Spatial Gradients

Labeling



$G_{x,y}$



Label-1

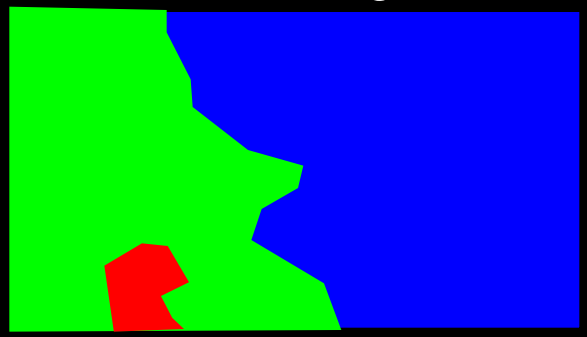
Label-2

Video frame

Inside a hole

Defining Spatial Gradients

Labeling



$G_{x,y}$

Label-1

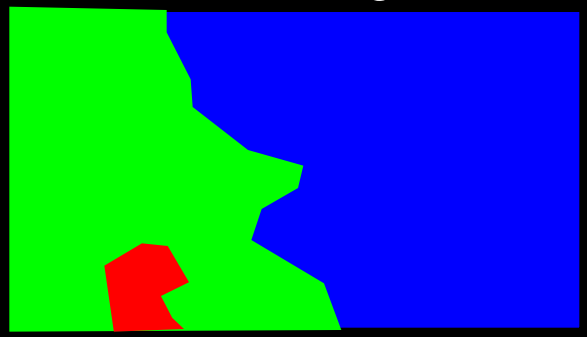
Label-2

Video frame

Inside a hole

Defining Spatial Gradients

Labeling



$G_{x,y}$

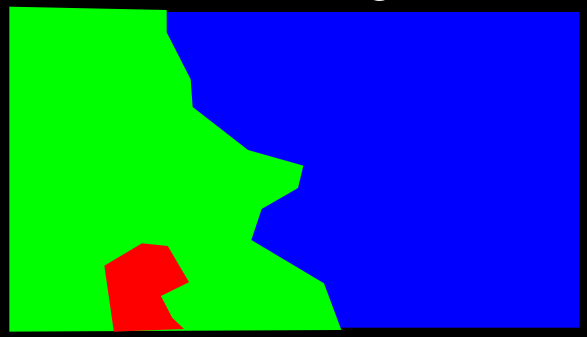
Label-1

Label-2

Video frame

Defining Spatial Gradients

Labeling



$G_{x,y}$

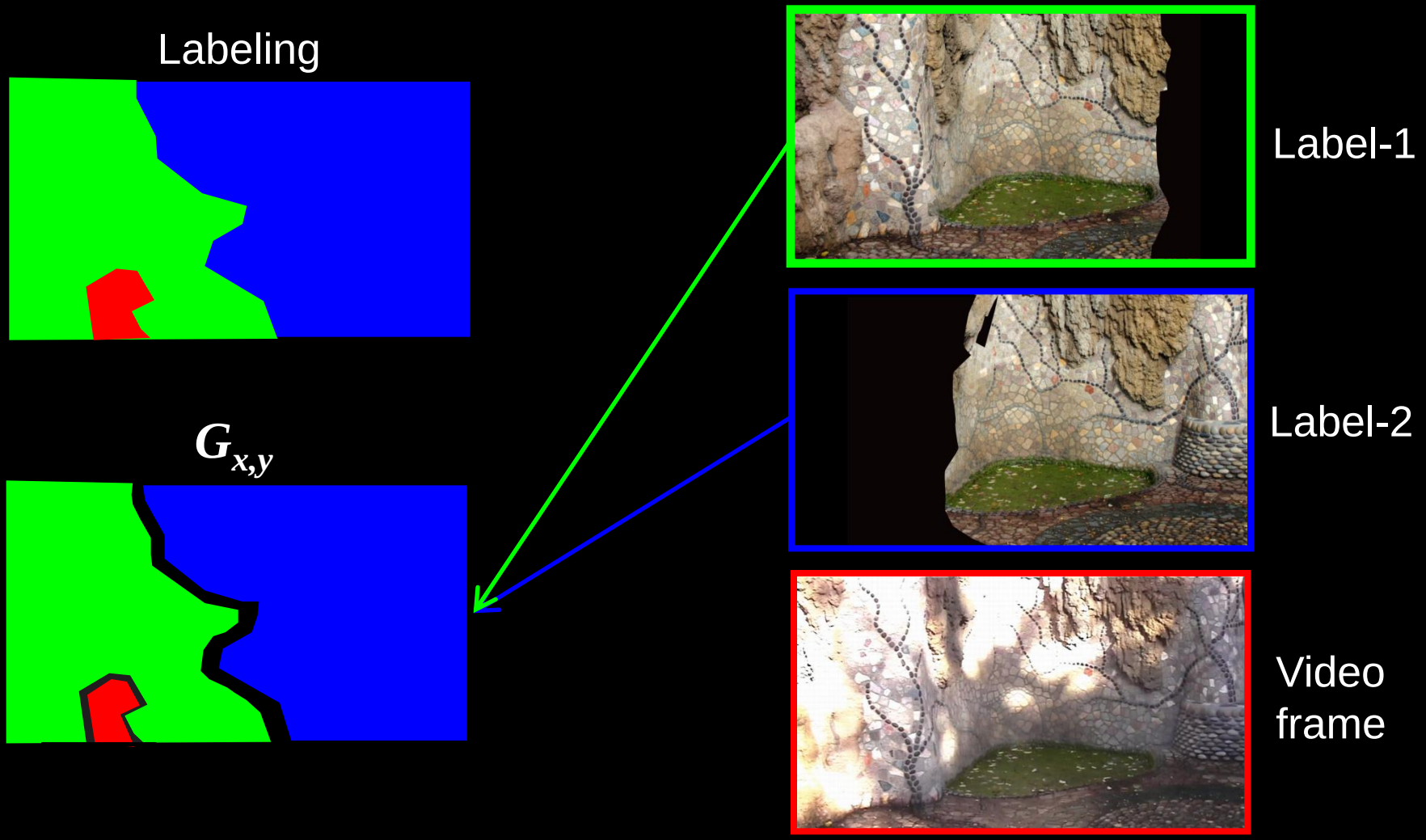
Label-1

Label-2

Video frame

Across seams

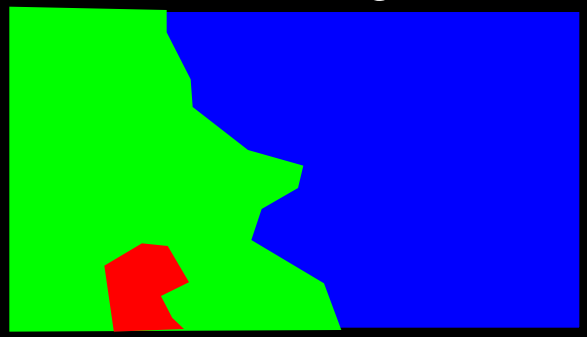
Defining Spatial Gradients



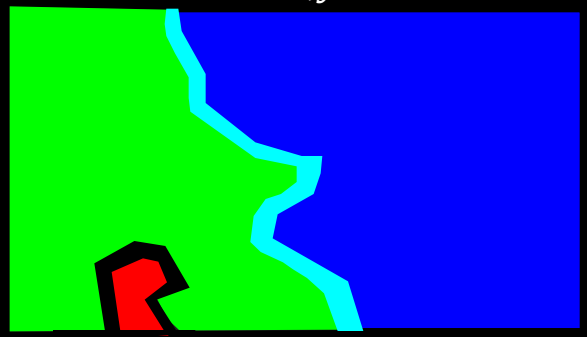
Across seams

Defining Spatial Gradients

Labeling



$G_{x,y}$



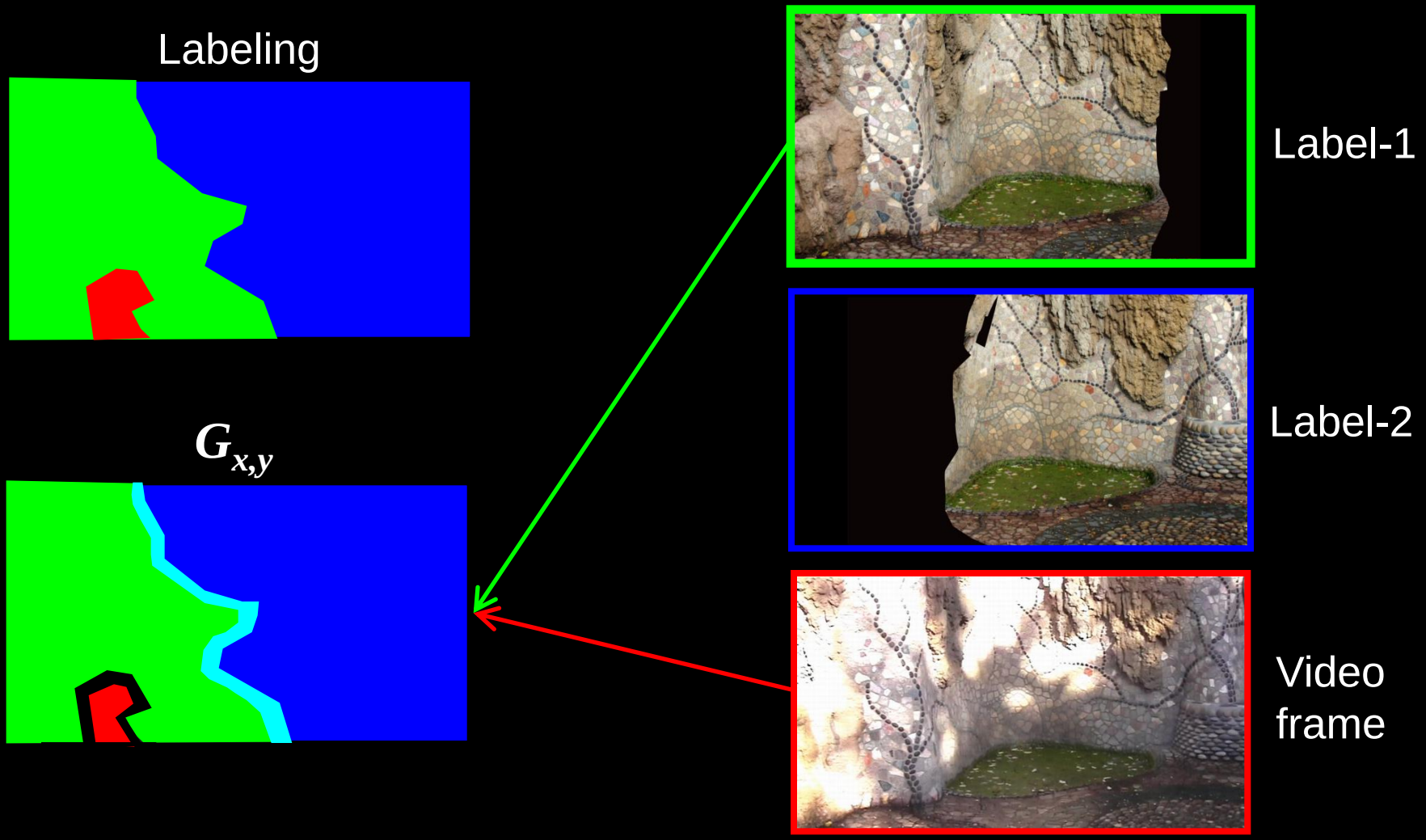
Label-1

Label-2

Video frame

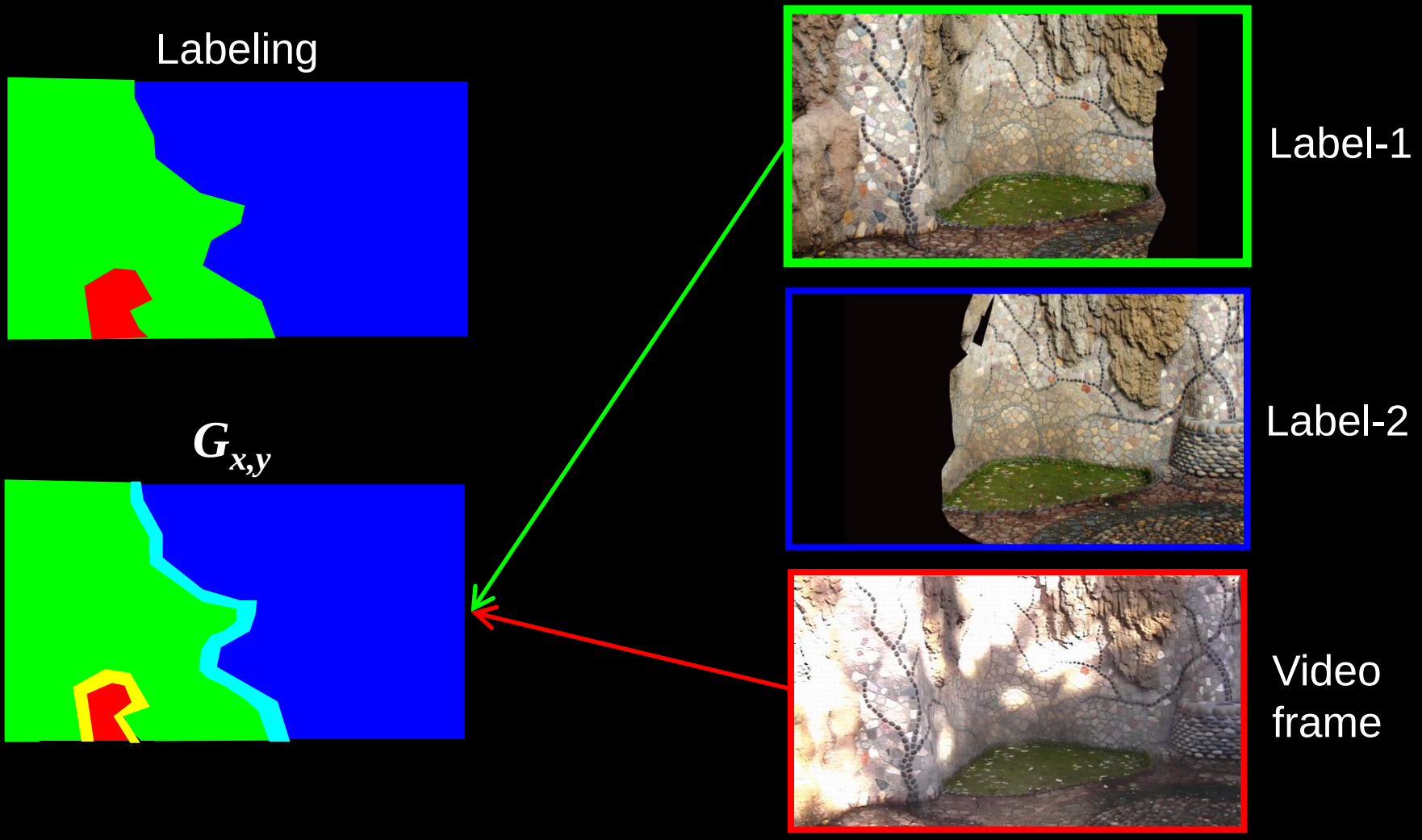
Across seams

Defining Spatial Gradients



Across seams

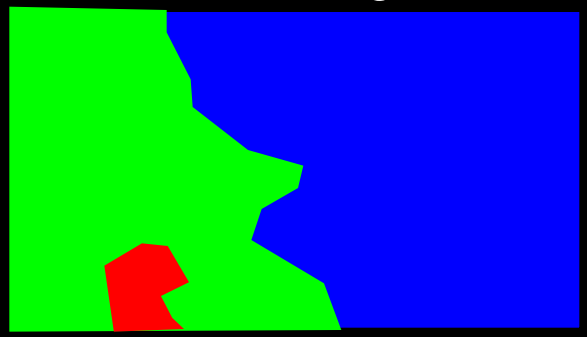
Defining Spatial Gradients



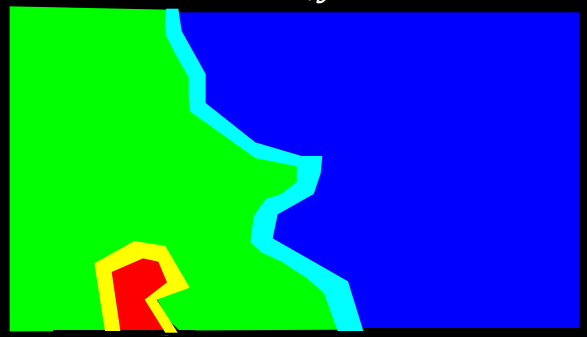
Across seams

Defining Spatial Gradients

Labeling



$G_{x,y}$



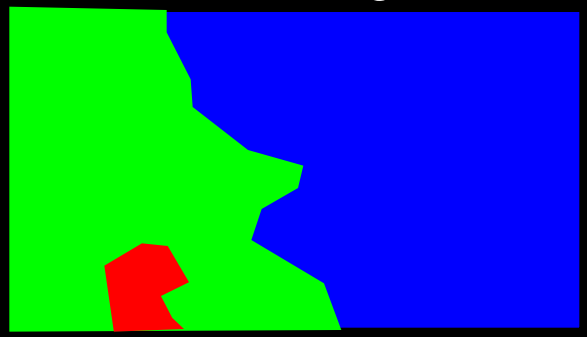
Label-1

Label-2

Video frame

Defining Spatial Gradients

Labeling



Enhanced Video Frame

Label-1

Label-2

Video frame

Integration result

Defining Spatial Gradients

Video reconstruction result

2D integration result

Spacetime Fusion

- 3D Gradient field of Enhanced Video

Spacetime Fusion

- 3D Gradient field of Enhanced Video
- Spatial gradients – G_x and G_y
  - as defined earlier

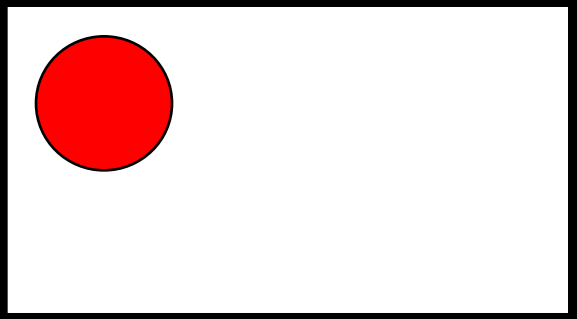
Spacetime Fusion

- 3D Gradient field of Enhanced Video
- Spatial gradients – G_x and G_y
  - as defined earlier
- Temporal gradients – G_t

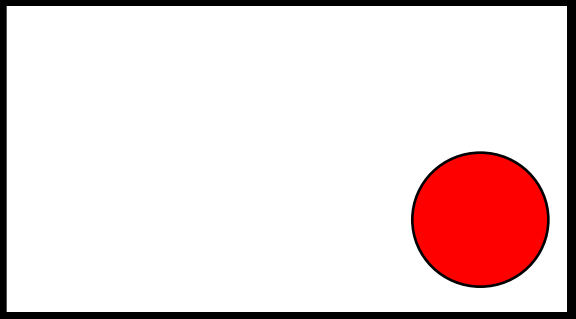
Spacetime Fusion

- 3D Gradient field of Enhanced Video
- Spatial gradients – $\boldsymbol{G}_x$ and $\boldsymbol{G}_y$
  - as defined earlier
- Temporal gradients – $\boldsymbol{G}_t$
  - $\boldsymbol{G}_t(x, y, t) = \boldsymbol{V}(x, y, t) - \boldsymbol{V}(x, y, t - 1)$

Spacetime Fusion

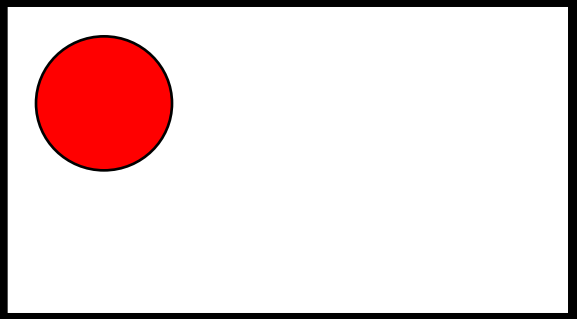


Video frame: t

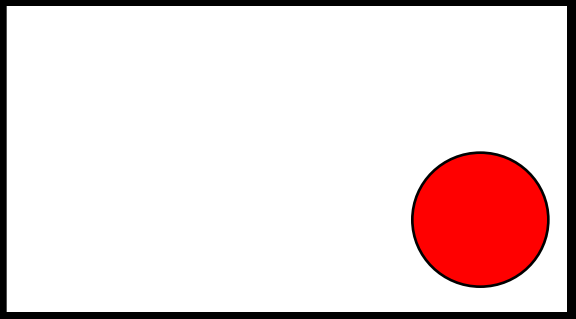


Video frame: t - 1

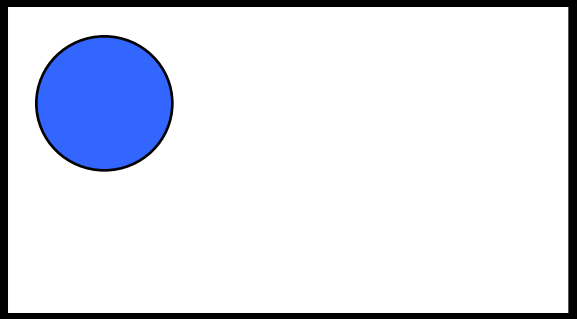
Spacetime Fusion



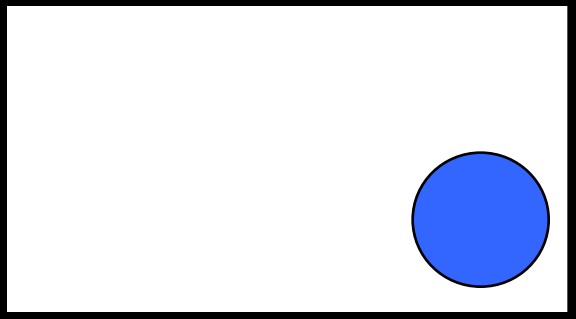
Video frame: t



Video frame: t - 1



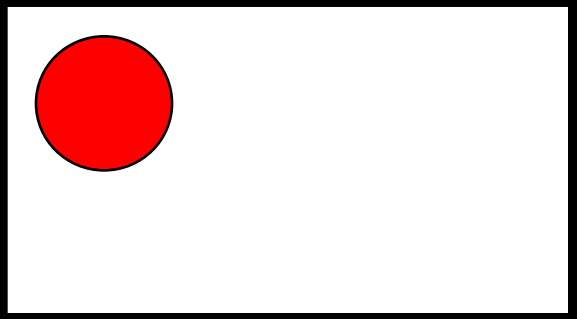
Reconstruction frame: t



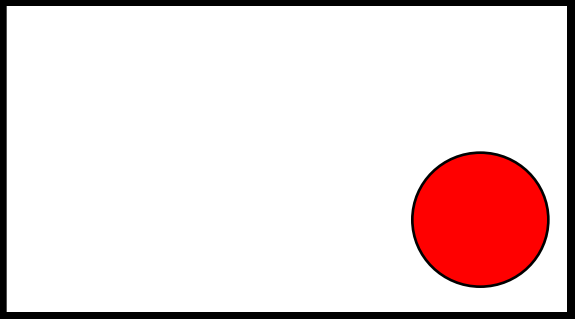
Reconstruction frame: t - 1

Spacetime Fusion

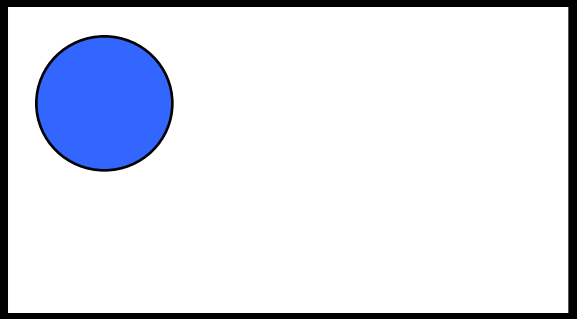
– $\boldsymbol{G}_t(x, y, t) = \boldsymbol{V}(x, y, t) - \boldsymbol{V}(x, y, t - 1)$



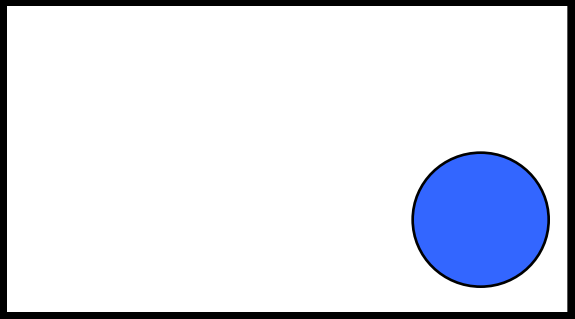
Video frame: t



Video frame: t - 1



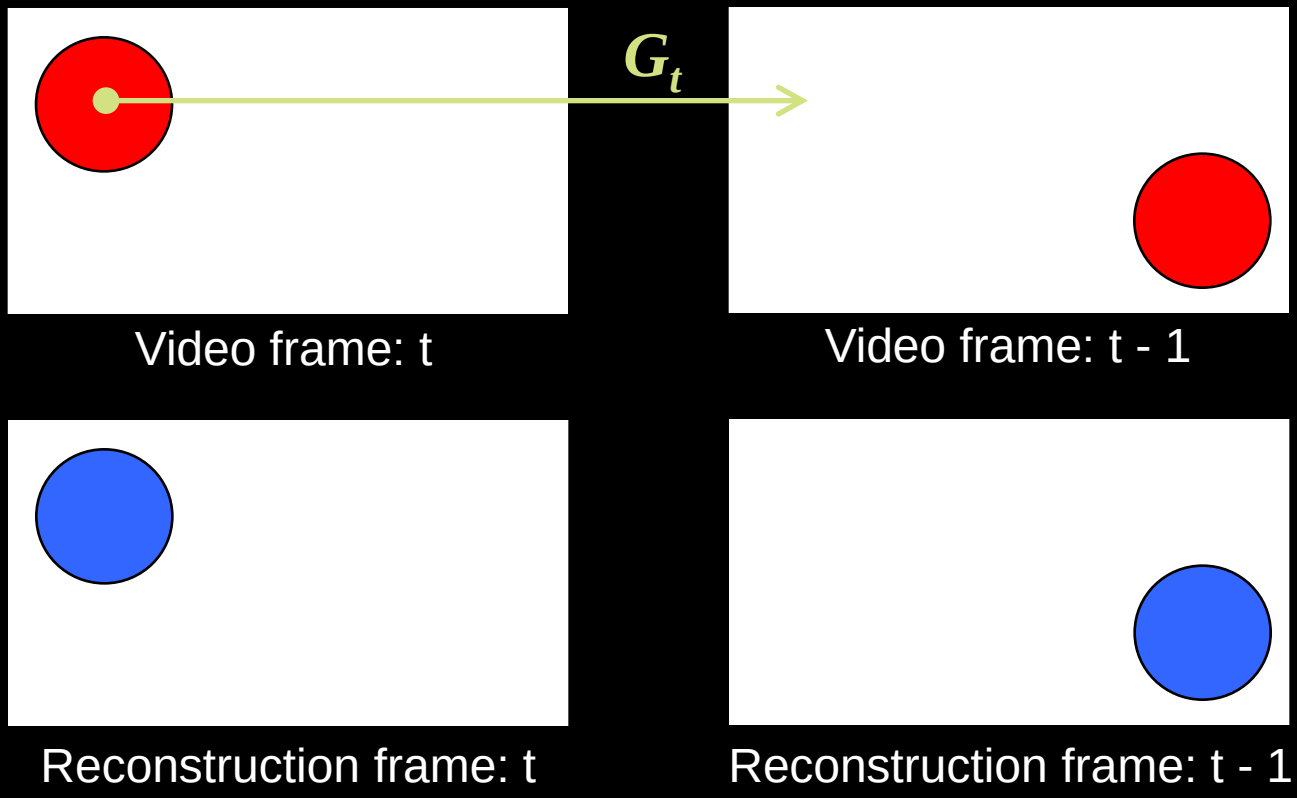
Reconstruction frame: t



Reconstruction frame: t - 1

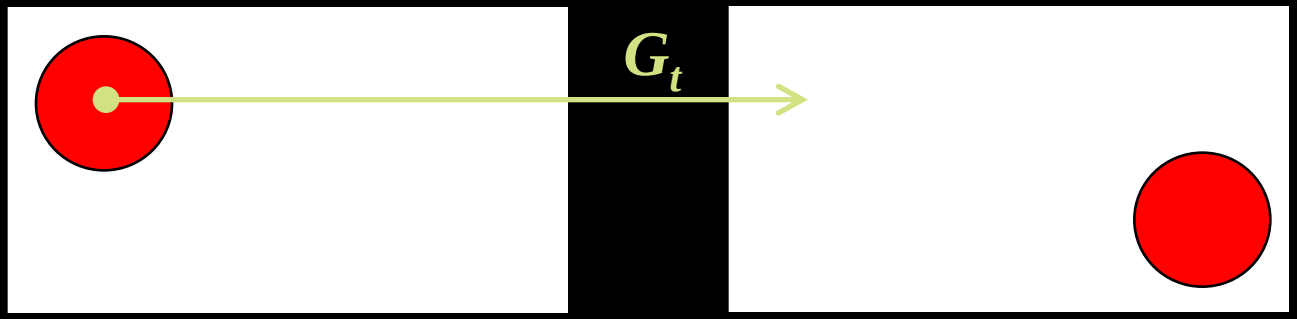
Spacetime Fusion

– $\boldsymbol{G_t}(x, y, t) = \boldsymbol{V}(x, y, t) - \boldsymbol{V}(x, y, t - 1)$



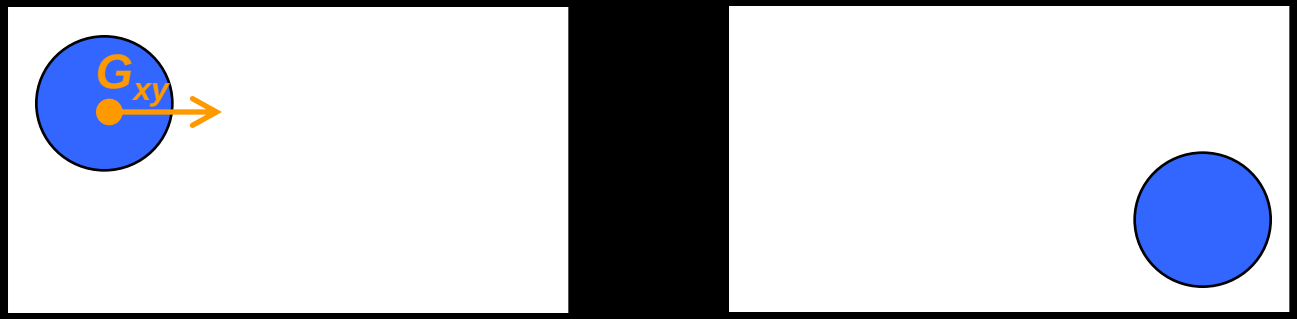
Spacetime Fusion

– $G_t(x, y, t) = V(x, y, t) - V(x, y, t - 1)$



Video frame: t

Video frame: t - 1

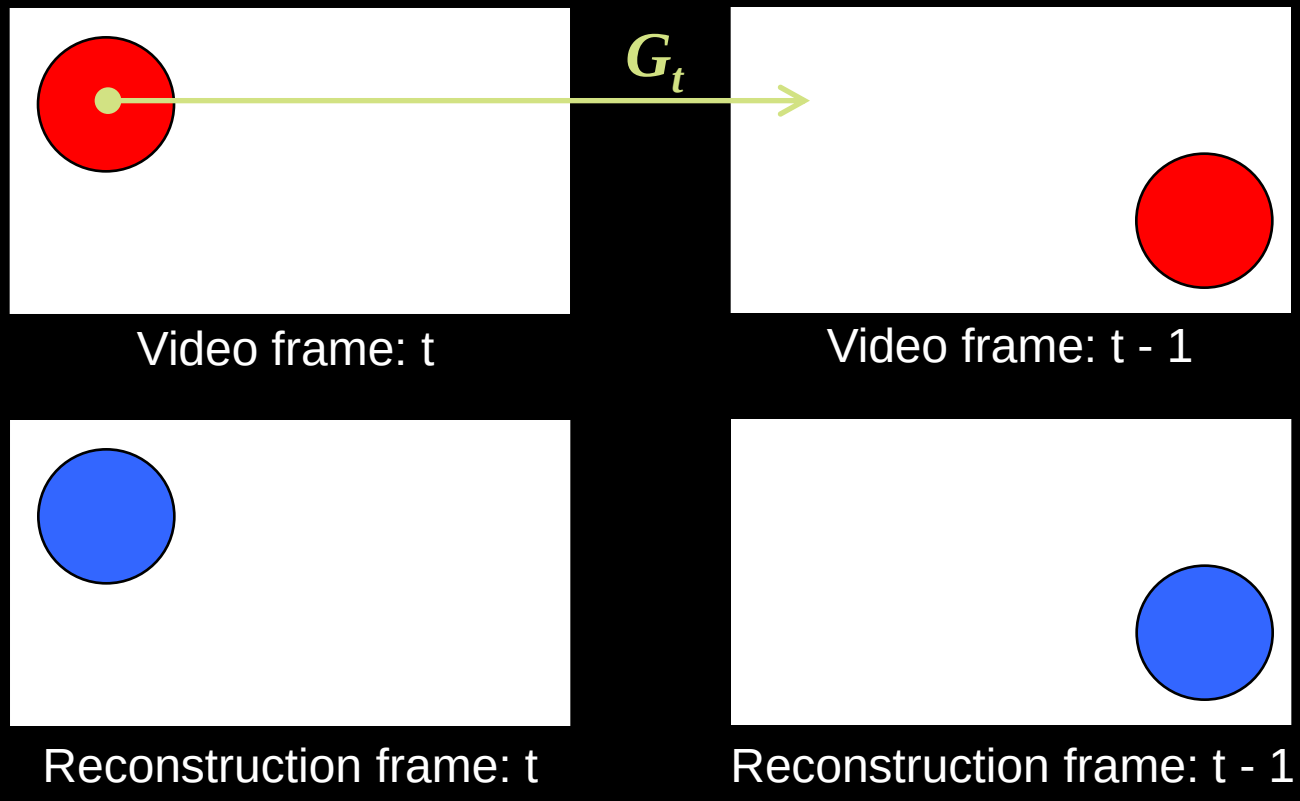


Reconstruction frame: t

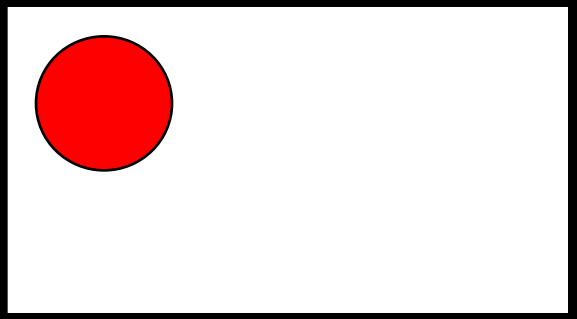
Reconstruction frame: t - 1

Spacetime Fusion

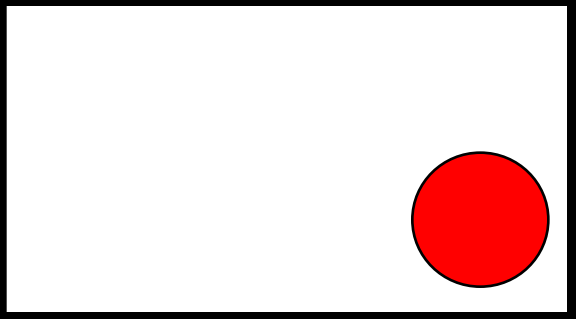
- G_t incompatible with $G_{x,y}$



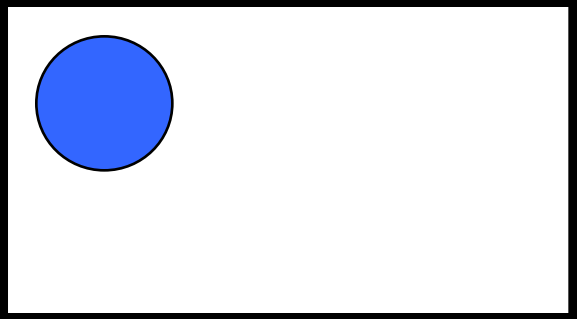
Spacetime Fusion



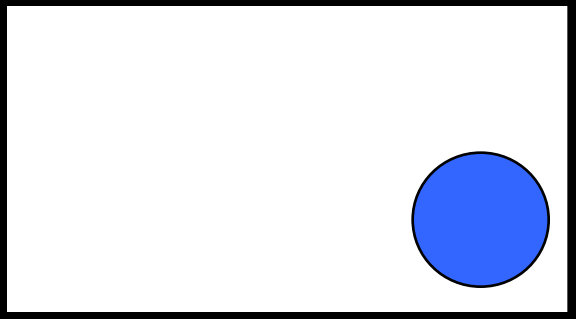
Video frame: t



Video frame: t - 1

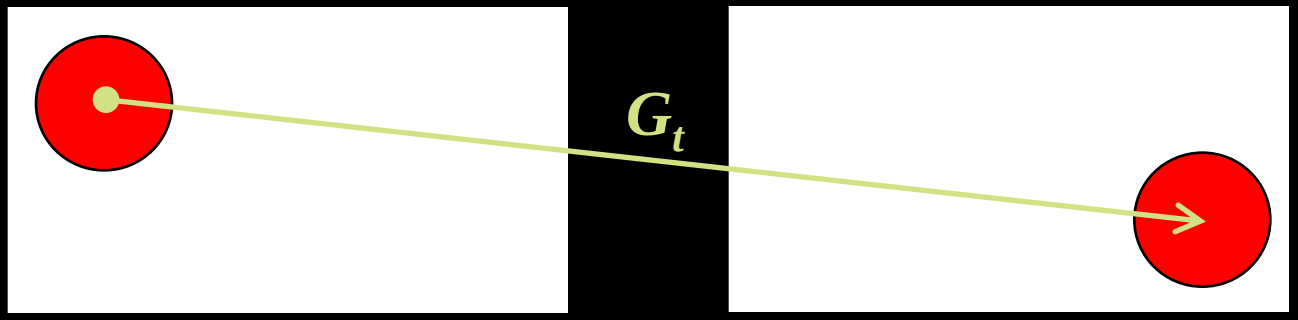


Reconstruction frame: t



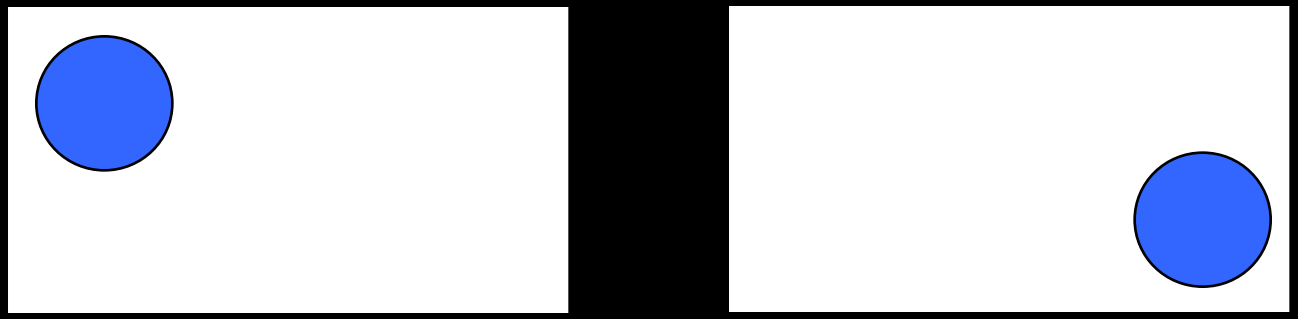
Reconstruction frame: t - 1

Spacetime Fusion



Video frame: t

Video frame: t - 1

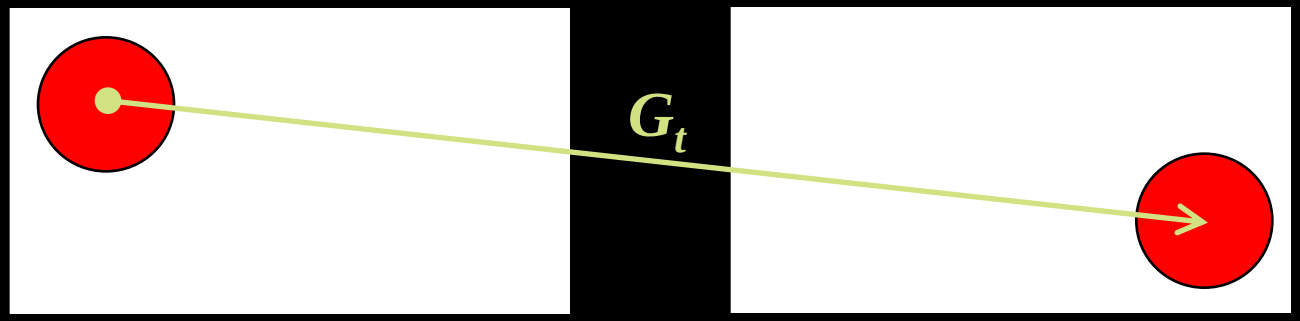


Reconstruction frame: t

Reconstruction frame: t - 1

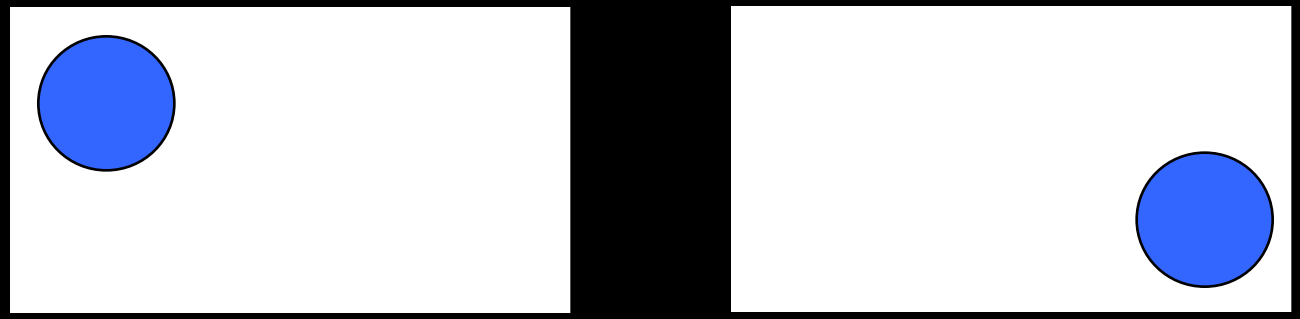
Spacetime Fusion

$x - u \quad y - v$



Video frame: t

Video frame: t - 1

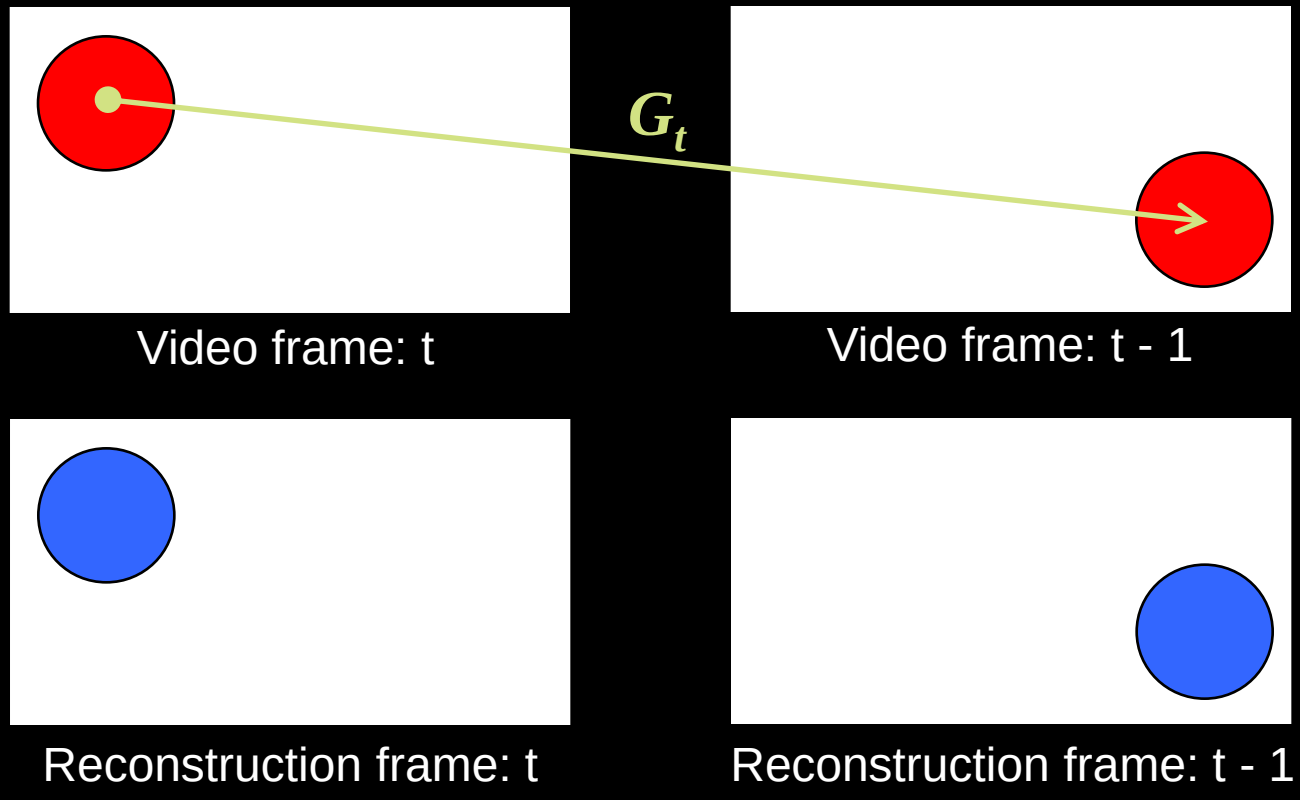


Reconstruction frame: t

Reconstruction frame: t - 1

Spacetime Fusion

- G_t compatible with $G_{x,y}$



Spacetime Fusion

- 3D Gradient field of Enhanced Video
- Spatial gradients – $\boldsymbol{G}_x$ and $\boldsymbol{G}_y$
  - as defined earlier
- Temporal gradients – $\boldsymbol{G}_t$
  - $\boldsymbol{G}_t(x, y, t) = \boldsymbol{V}(x, y, t) - \boldsymbol{V}(x, y, t - 1)$

Spacetime Fusion

- 3D Gradient field of Enhanced Video
- Spatial gradients – $\boldsymbol{G_x}$ and $\boldsymbol{G_y}$
  - as defined earlier
- Temporal gradients – $\boldsymbol{G_t}$
  - $\boldsymbol{G_t}(x, y, t) = \boldsymbol{V}(x, y, t) - \boldsymbol{V}(x - u \;\; y - v, t - 1)$

Where (u,v) is the correspondence vector

Spacetime Fusion

- Solving for Enhanced Video

Solve linear system:
$\boldsymbol{Av} = \boldsymbol{b}$

Spacetime Fusion

- Solving for Enhanced Video

Solve linear system:

$\boldsymbol{Av} = \boldsymbol{b}$

Constraints:

$\boldsymbol{V}_{x, y, t} - \boldsymbol{V}_{x-1, y, t} = \boldsymbol{G}_{x}(x, y, t)$

$\boldsymbol{V}_{x, y, t} - \boldsymbol{V}_{x, y-1, t} = \boldsymbol{G}_{y}(x, y, t)$

$\boldsymbol{V}_{x, y, t} - \boldsymbol{V}_{x-u, y-v, t} = \boldsymbol{G}_{t}(x, y, t)$

Spacetime Fusion

Video reconstruction result

Spacetime fusion result

Applications

Enhanced Exposure

Super-Resolution

Input images

Video Editing

Edit Propagation

Edit Propagation

- User Interaction
  - User edits one or more images

Edit Propagation

- User Interaction
  - User edits one or more images
- Solution
  - Reconstruct video frames using edited pixels

Edit Propagation

Video Editing

Input video frame

Edit Propagation

Input Video

Edit Propagation

User Edits

Edit Propagation

User Edits

Edit Propagation

User Edits

Edit Propagation

User Edits

Edit Propagation

User Edits

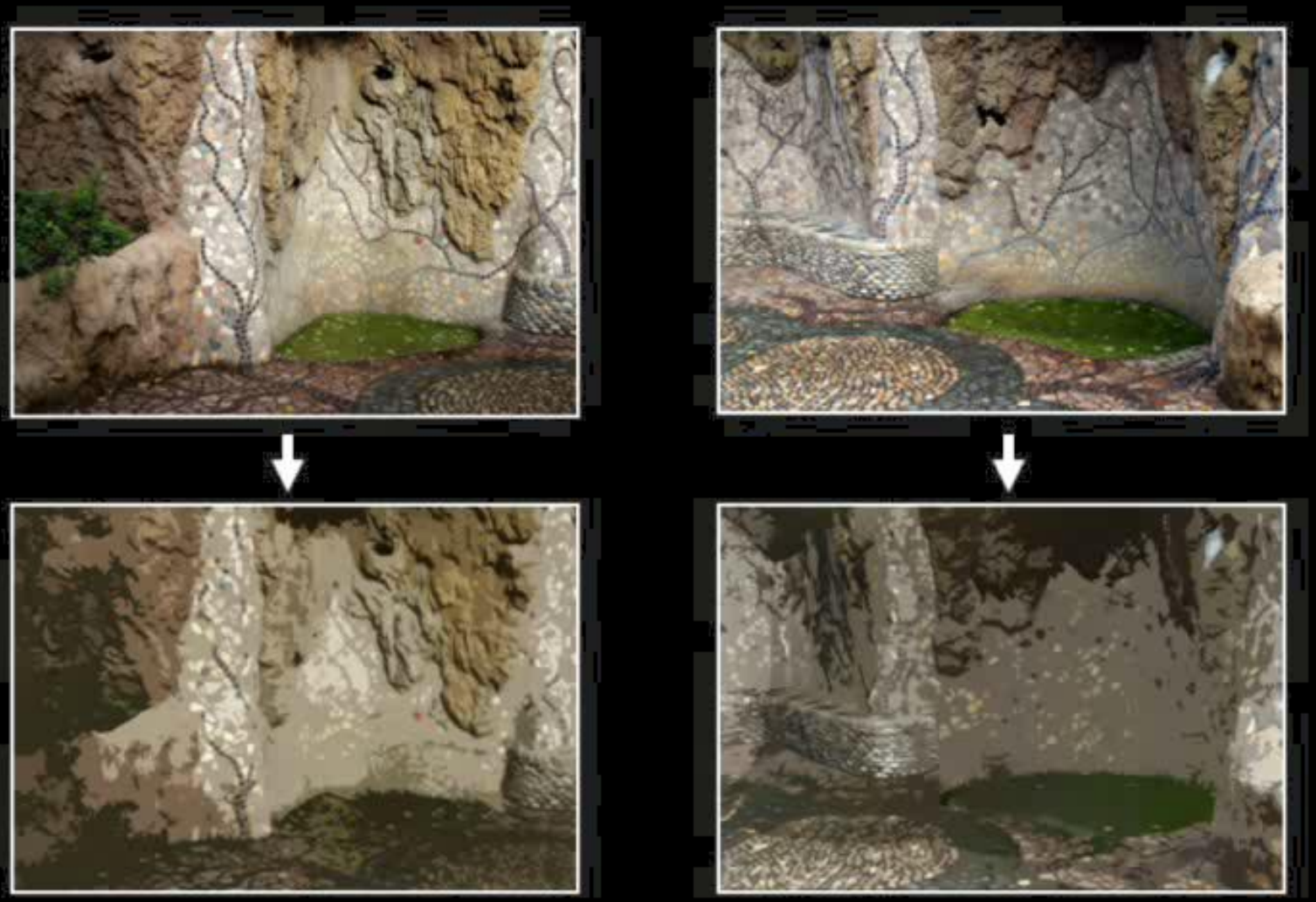
Edit Propagation

User Edits

Edit Propagation

Edited Video

Edit Propagation



Painterly effects

Camera Shake Removal

Camera Shake Removal

- Solution
  - Smooth out input-video camera path
  - Re-render video from new camera path

Camera Shake Removal

Input Video

Camera Shake Removal

Stabilized Video

Object Removal

Object Removal

- User Interaction
  - User draws object mask in one image

Object Removal

- User Interaction
  - User draws object mask in one image
- Solution
  - Transfer mask to all video frames
  - Reconstruct masked portions using unmasked portions

Object Removal

Original video

Conclusion

Conclusion

- General framework
- Enhanced multi-view stereo
- Novel IBR
  - Renders dynamic lighting
  - Avoids ghosting
  - Temporarily consistent
- IBR framework can be applied to dynamic scenes

Future Work

- Speed ups
- UI
- Dynamic scenes

Thanks

- Reviewers / EGSR organizers
- Advisors - Michael Cohen and Brian Curless
- NVIDIA Fellowship

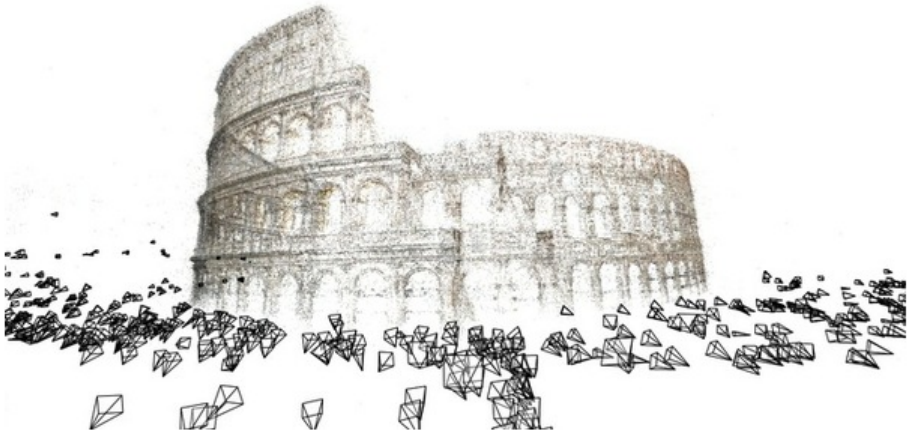
Performance

- 5 Minutes per frame (853x480)
  - 2 minutes on SFM
  - 2 minutes on MVS
  - 1 minute on IBR

Photo Count

- Grotto – 11 photos
- Bust – 7 photos
- Dessert table – 12 photos
- Suzzalo – 6 photos

Computers Turn Flat Photos Into 3-D Buildings



GRAND SCALE A 3-D reconstruction of the Colosseum in Rome, built as part of the "Rome in a Day" project, which used 2,106 images and 819,242 points.

By JOHN MARKOFF
Published: February 22, 2010

Rome wasn't built in a day, but in cyberspace it might be.

Computer science researchers at the University of Washington and Cornell

SIGN IN TO RECOMMEND

TWITTER

SIGN IN TO E-MAIL

RSS Feed

Next Article in Science (7 of 32) »

“Photo Tourism” a.k.a. Photosynth

Siggraph 2006

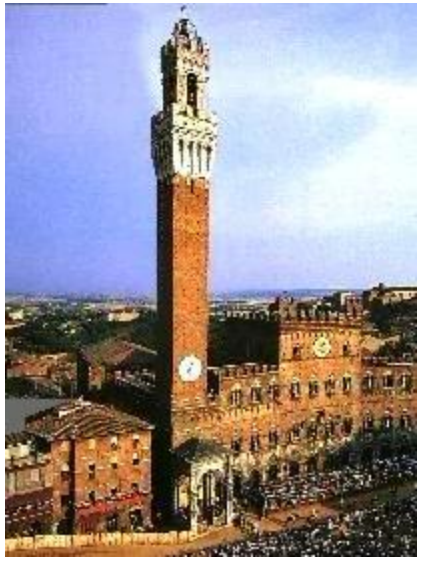
Applet Photosynth on tv: “CSI” Windmill Photosynth

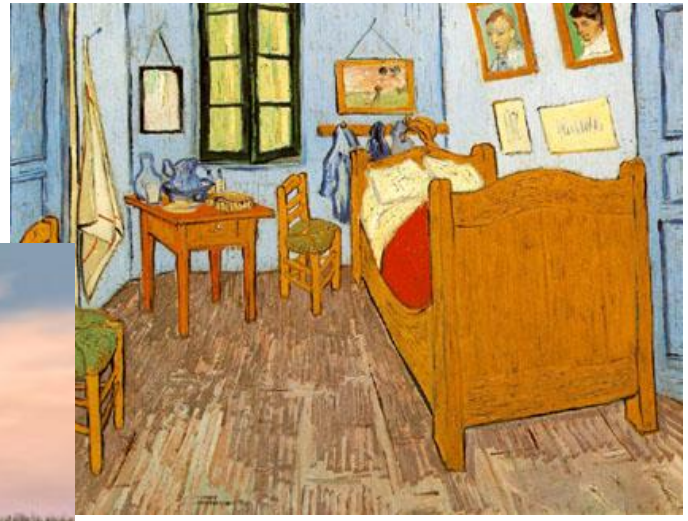
See more on reconstruction from Community Photo Collections on Noah Snavely’s or the University of Washington web pages.

See also:
User in the Loop?

Canoma (was a product…)

From http://www.canoma.com/movies.html





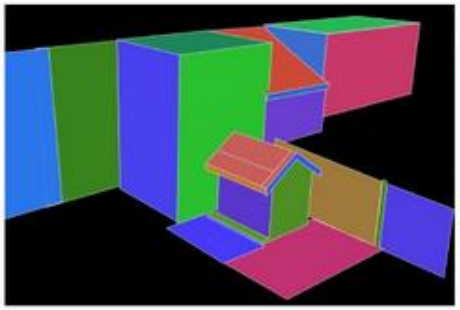
"Beached Canoma" - Image by Don Gray

See also…

- Interactive 3D Architectural Modeling from Unordered Photo Collections
  - Sinha et al., SIGGRAPH ASIA 2008
  - Project webpage

(a) Input Photographs

(b) 2D Sketching Interface



(c) Geometric Model

(d) Texture-mapped model

- Interactive 3D
  Unordered Ph
  - Sinha et al.
  - Project we

(a) Input Photographs (b)

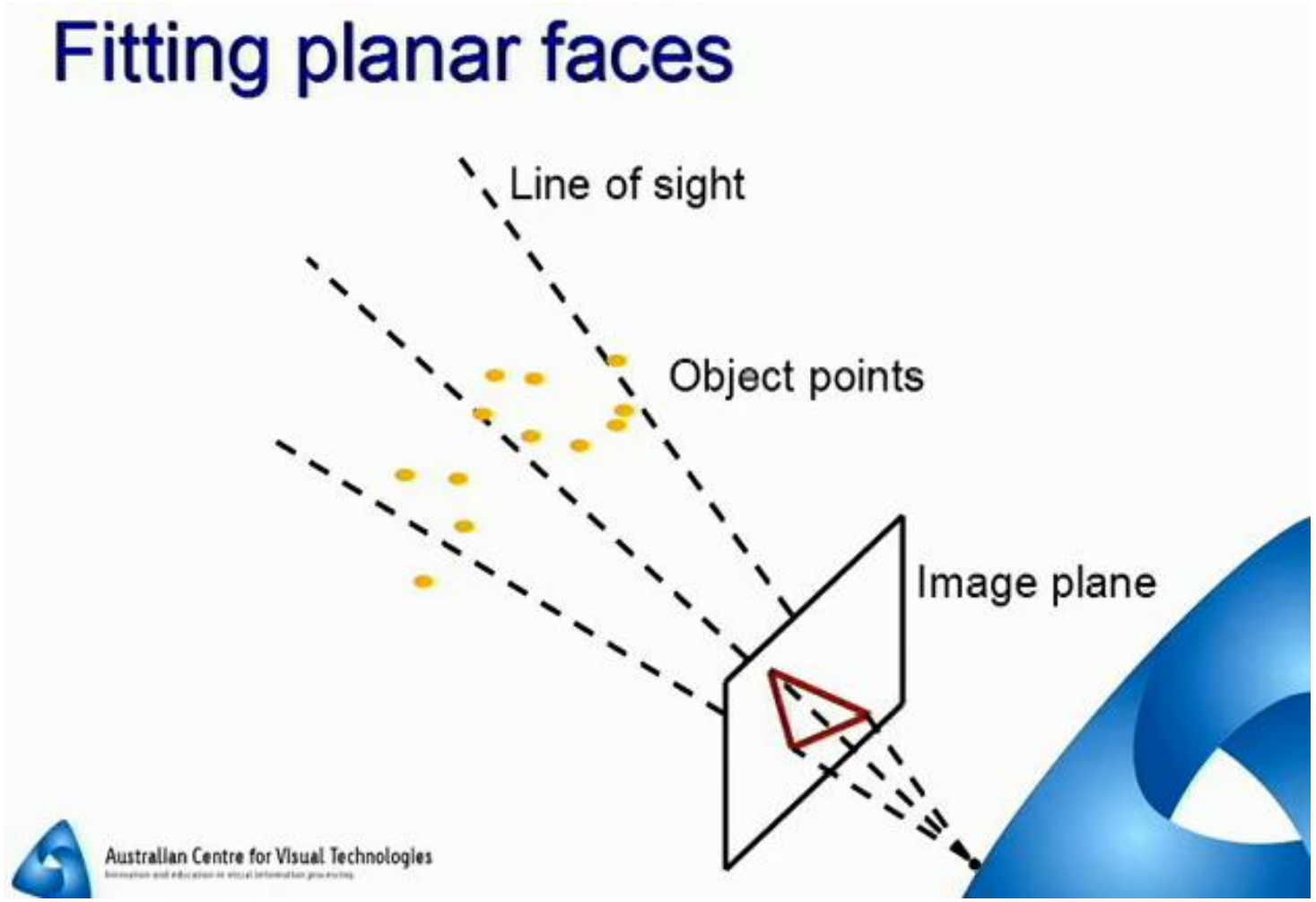
VideoTrace: Rapid interactive scene modelling from video, SIGGRAPH 2007

- By van den Hengel, Dick, Thormählen, Ward, Torr
- Project Website, GoogleTalk
- Being commercialized
- See also: free structure-from-motion package called Voodoo webpage





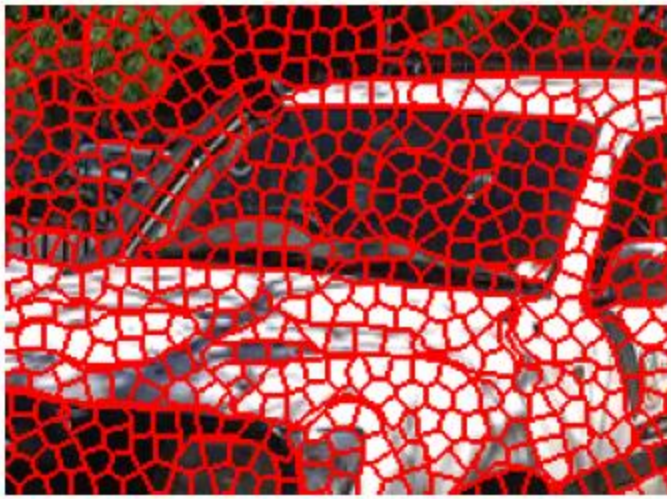

- User helps model object as it appears in the video
  - Not necessarily accurate
  - Not necessarily reusable



- RANSAC fitting of planes
  - based on color histograms, color-matching of edge projections, check for self-occlusions

From Anton van den Hengel



Dense meshing is still a big problem!