

Optimal Representation of Multi-View Video

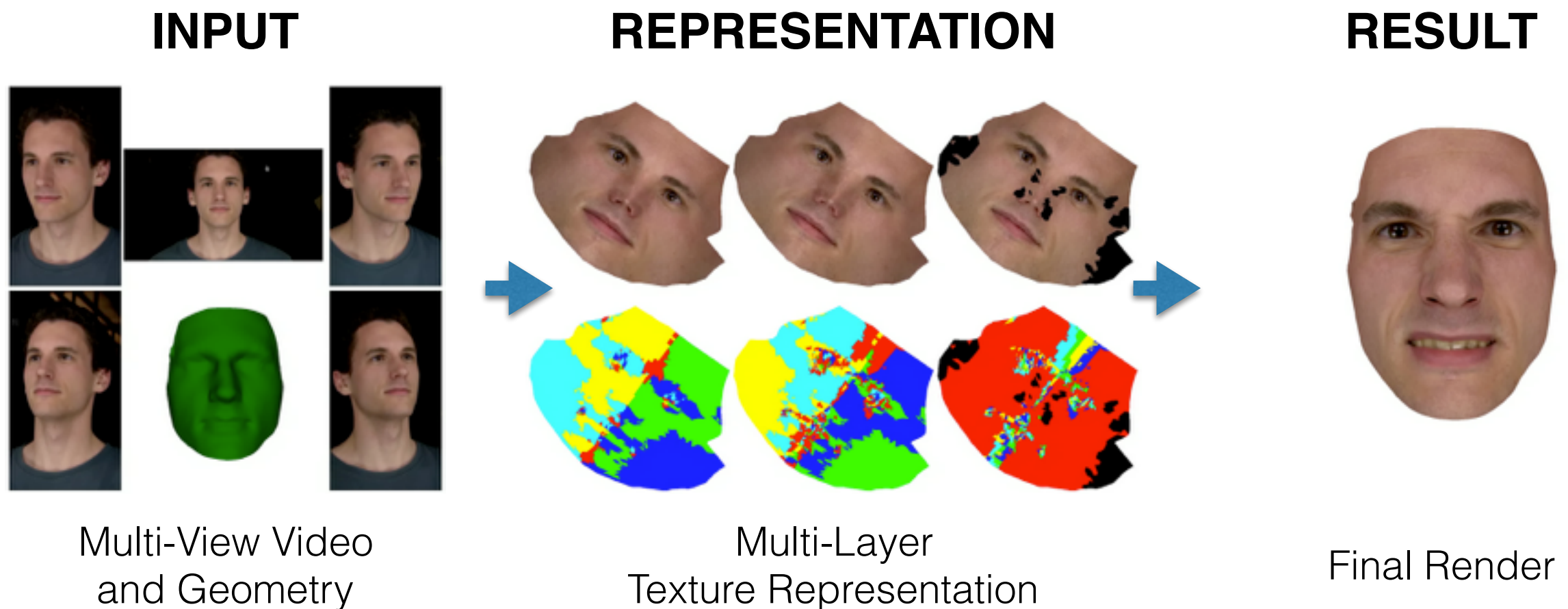
M. Volino, D. Casas, J. Collomosse, A. Hilton

BRITISH MACHINE VISION CONFERENCE 2014



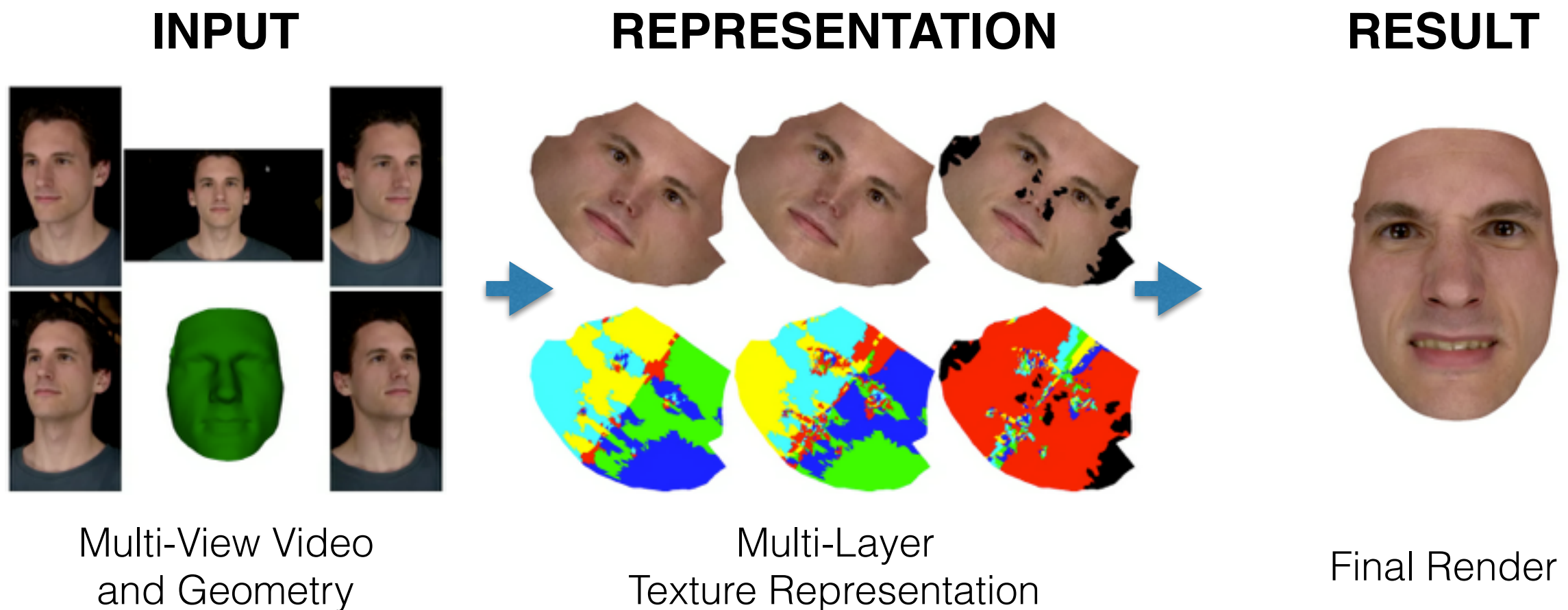
Problem Statement

Optimal resampling of the captured views to obtain a compact representation without loss of view-dependent dynamic surface appearance information



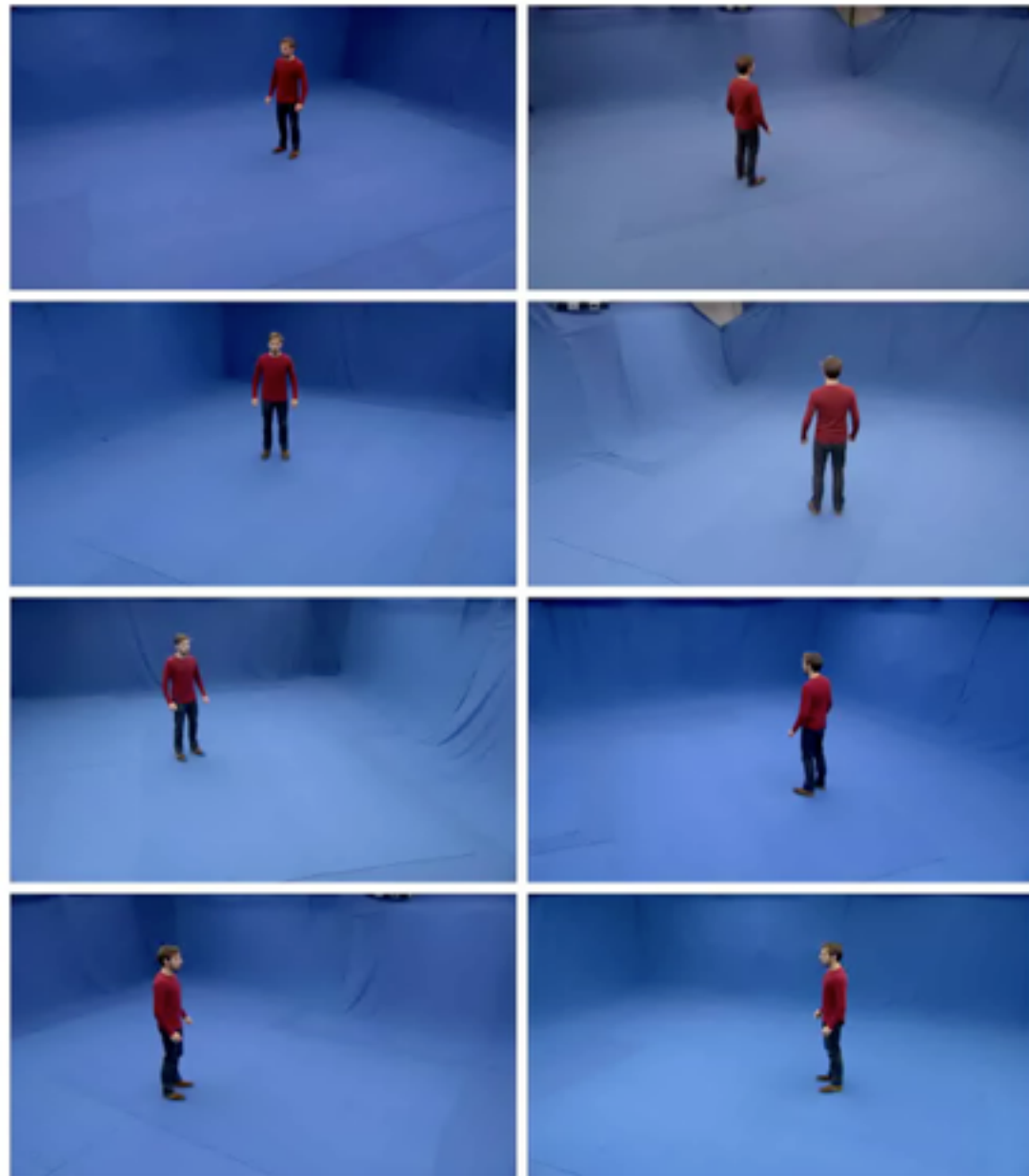
Problem Statement

Optimal resampling of the captured views to obtain a compact representation without loss of view-dependent dynamic surface appearance information



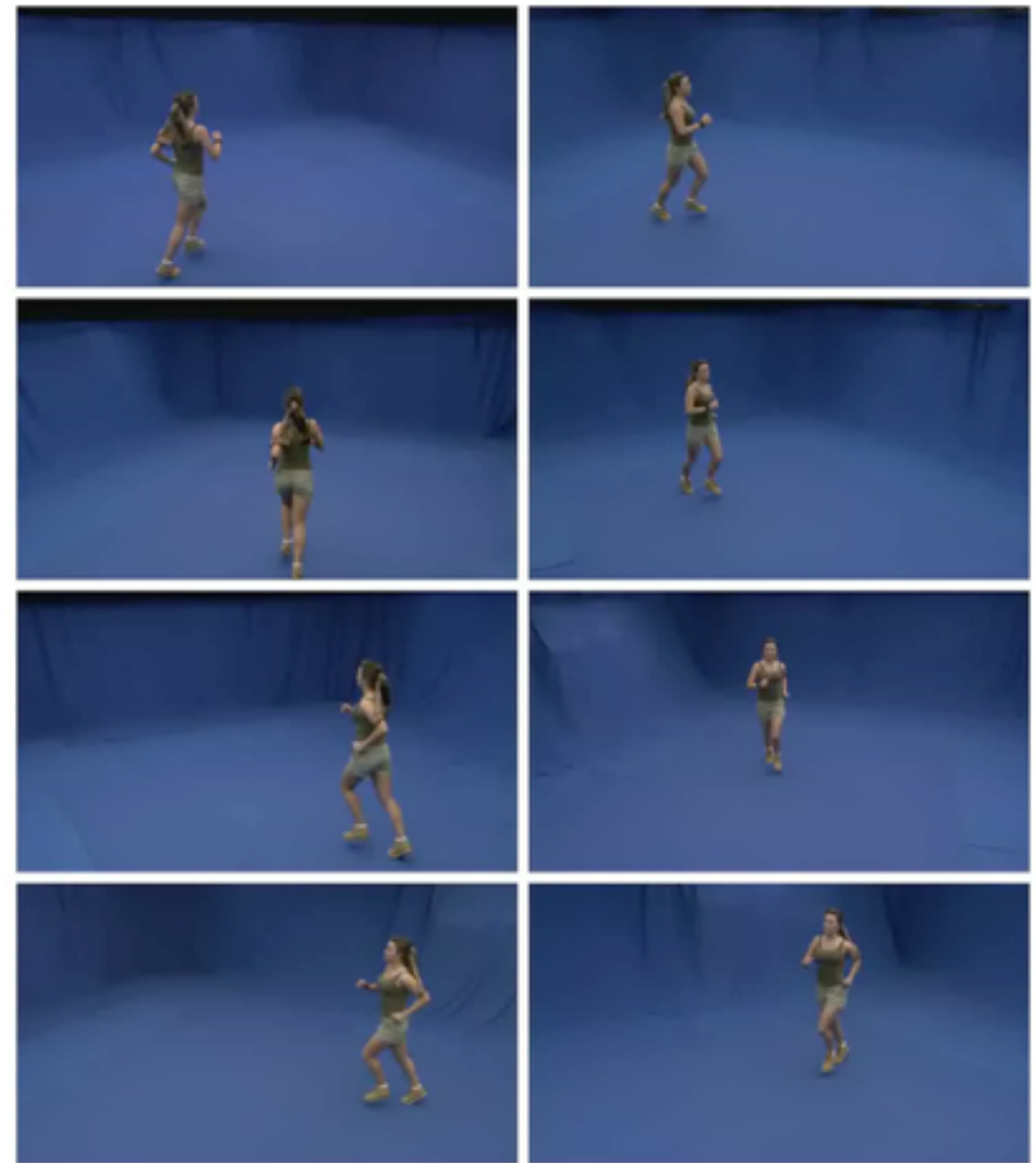
Multi-View Video

Character Dan



Storage: 1.6 GB

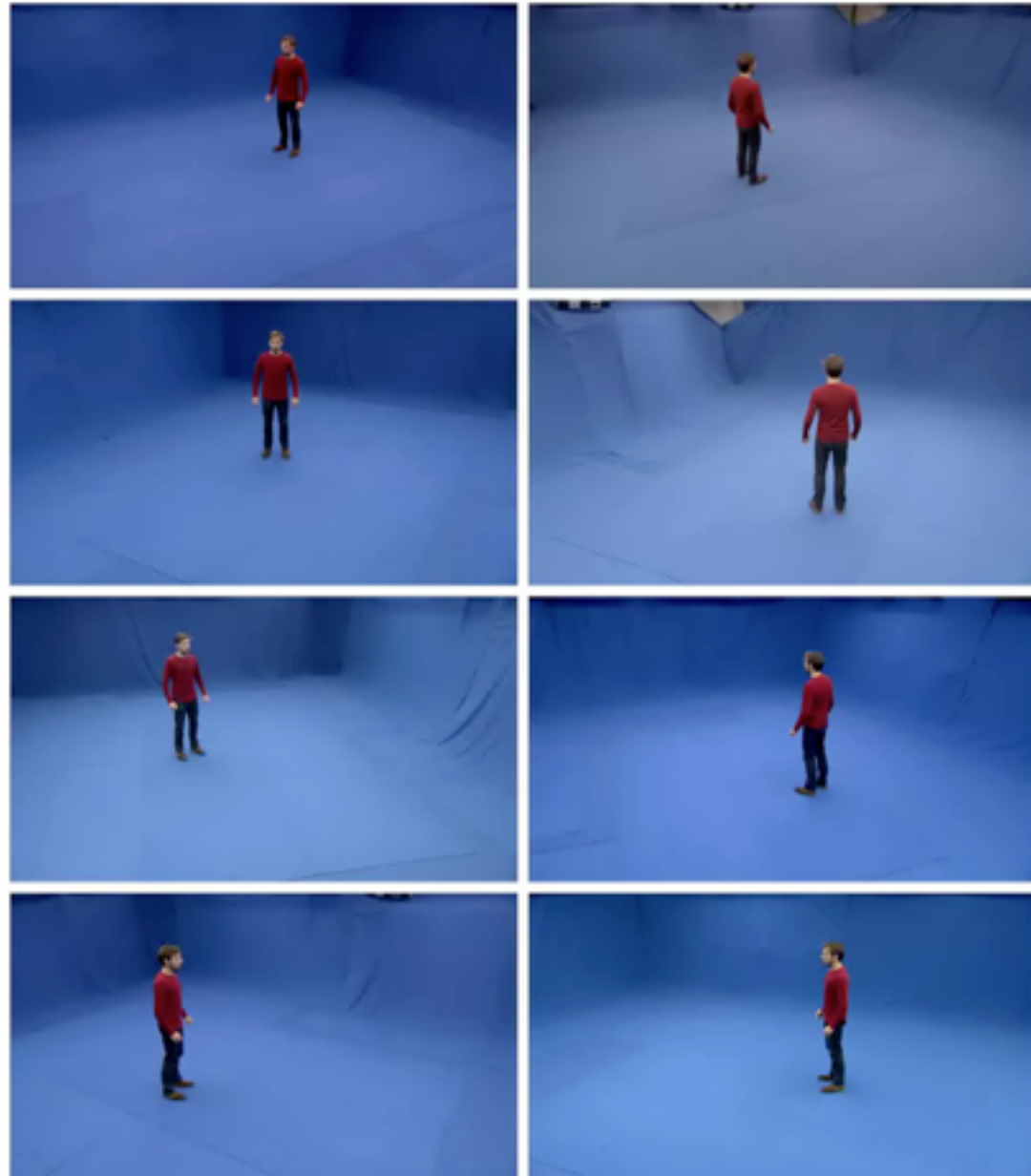
Character 1



Storage: 1.8 GB

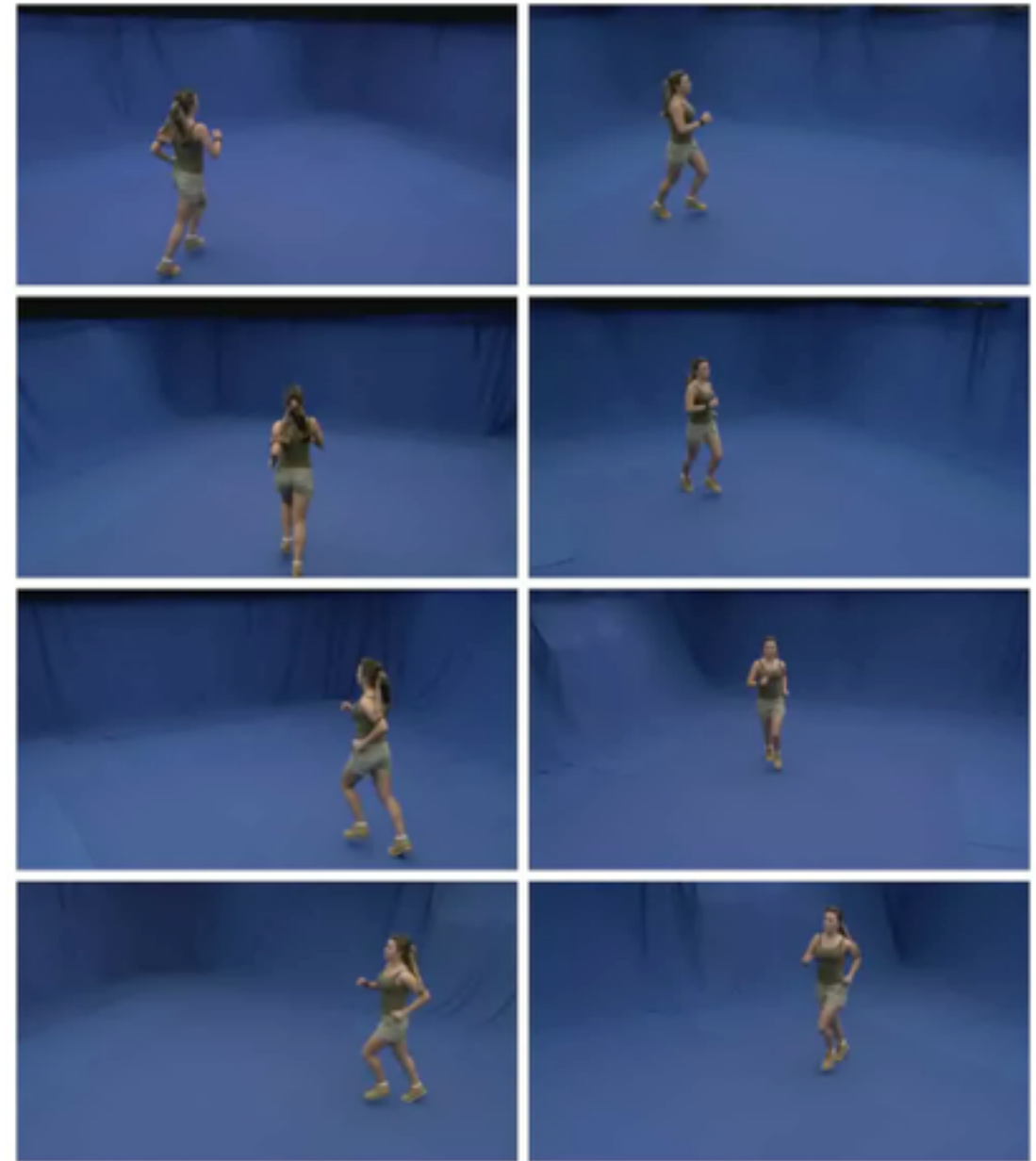
Multi-View Video

Character Dan



Storage: 1.6 GB

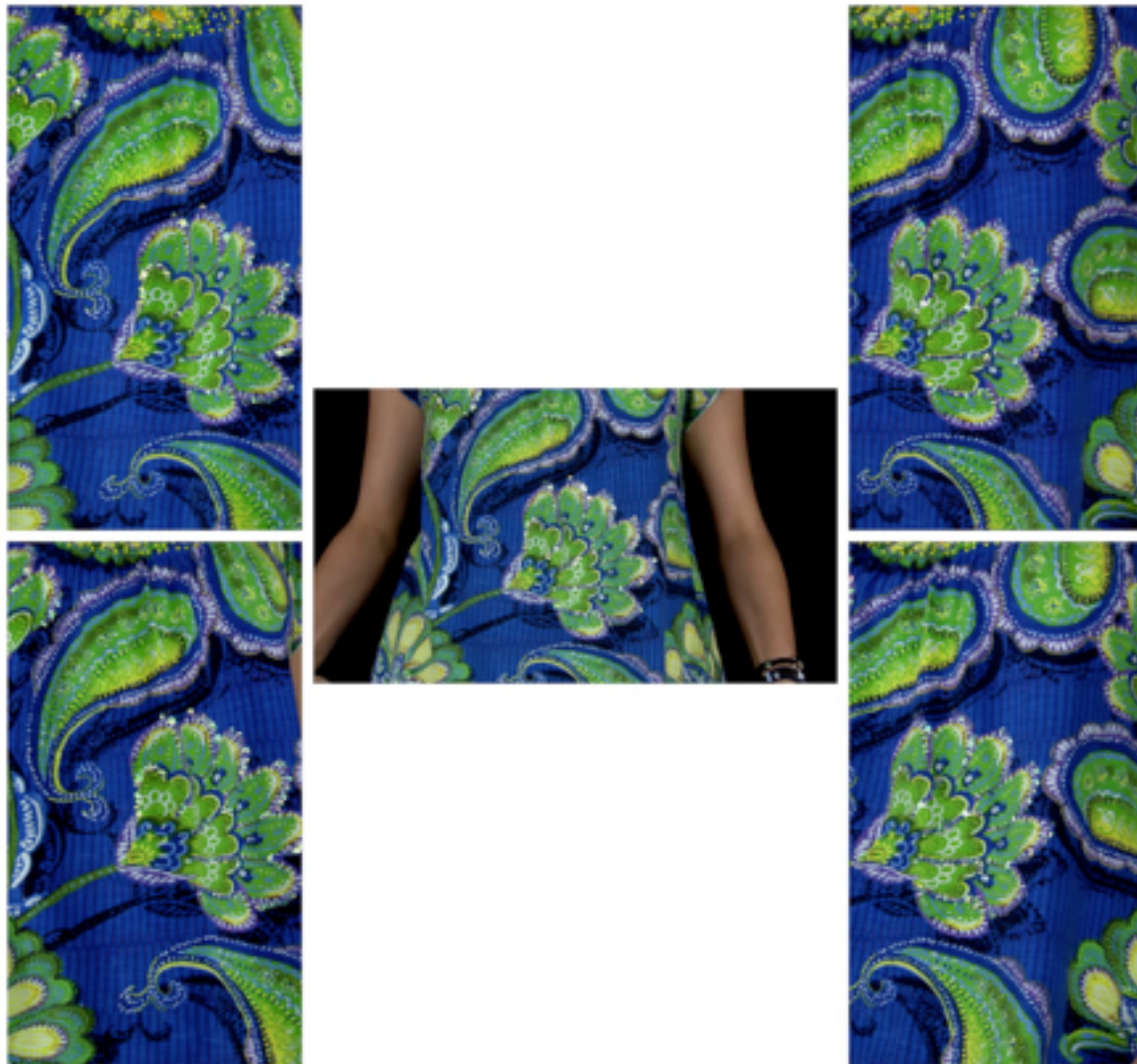
Character 1



Storage: 1.8 GB

Multi-View Video

Cloth



Storage: 11.4 GB

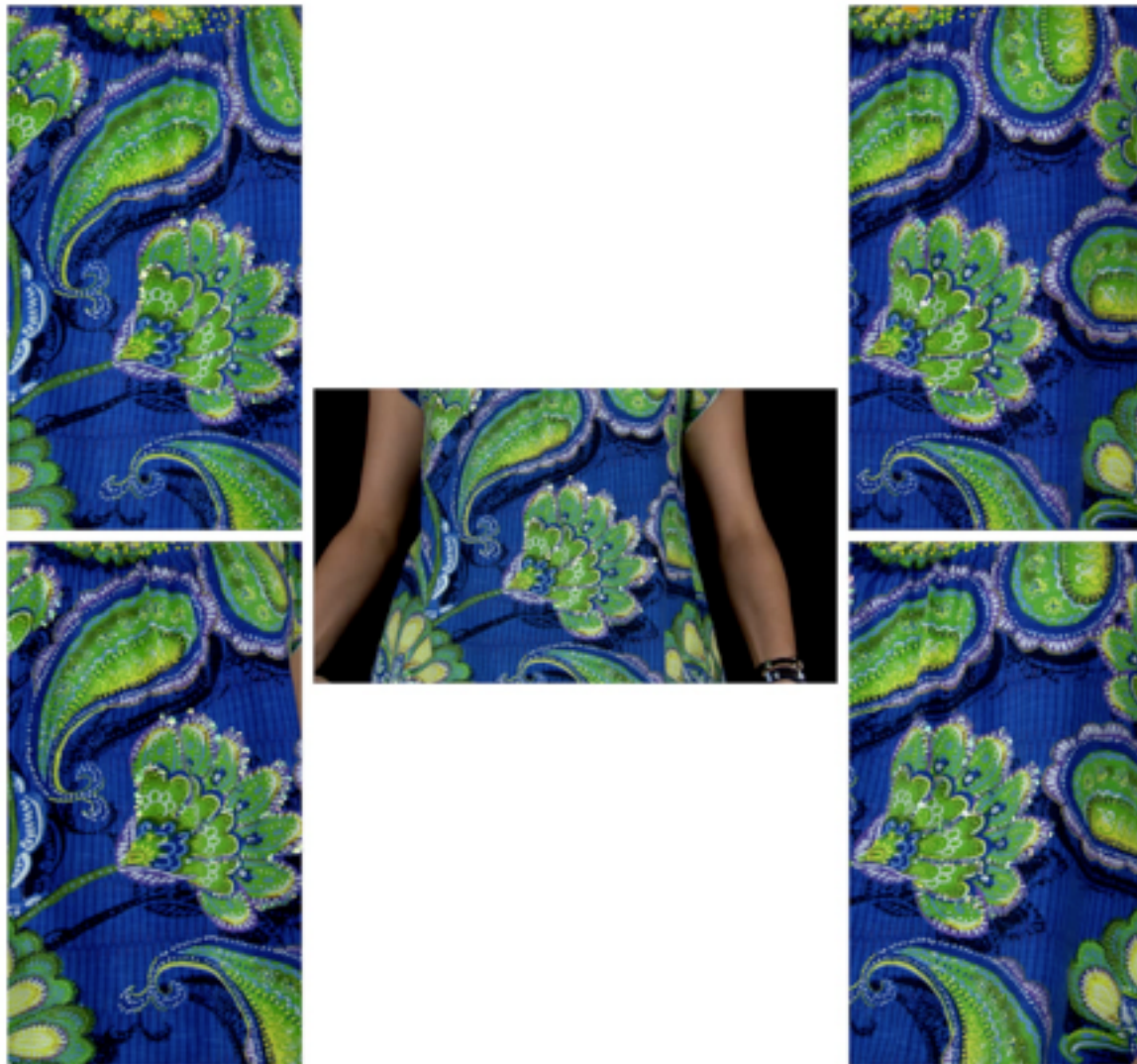
Face



Storage: 13.1 GB

Multi-View Video

Cloth



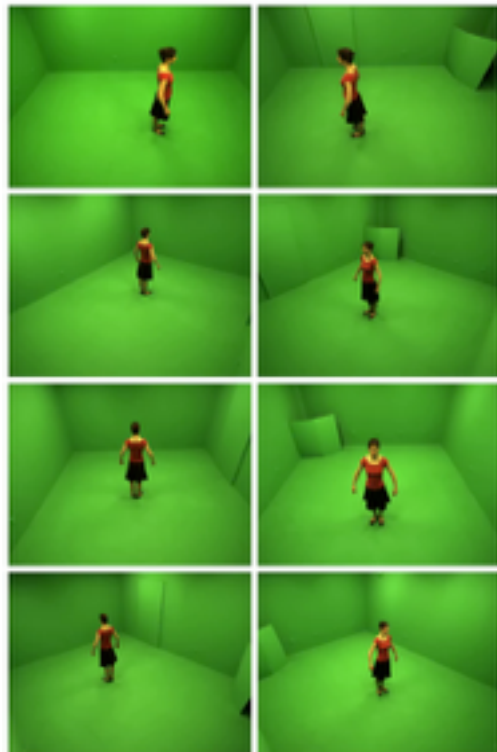
Storage: 11.4 GB

Face



Storage: 13.1 GB

Reconstructing Geometry



Vlasic 2008

Reconstructing Geometry

Method

Advantage

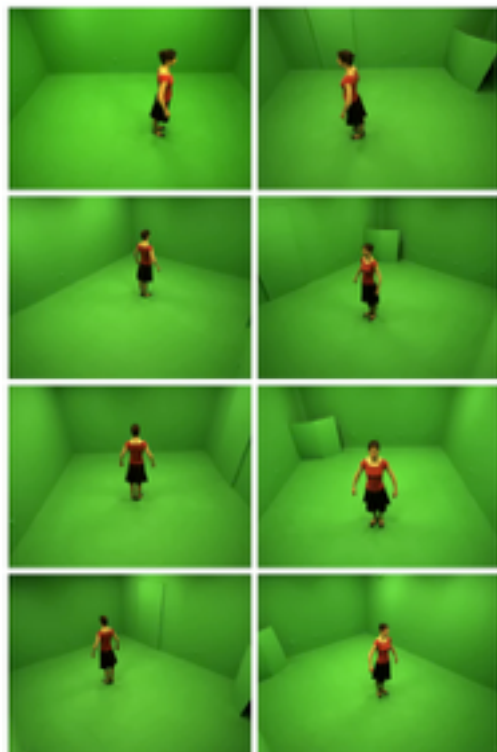
Disadvantage

Model-Based

- Highly detailed geometry
- Temporal consistent geometry

- Requires prior knowledge of subject
- Loss of dynamic surface detail
- Expensive equipment

Vlasic 2008
Carranza 2003



Vlasic 2008



Model-Based - Vlasic 2008

Reconstructing Geometry

Method

Advantage

Disadvantage

Model-Based

- Highly detailed geometry
- Temporal consistent geometry

- Requires prior knowledge of subject
- Loss of dynamic surface detail
- Expensive equipment

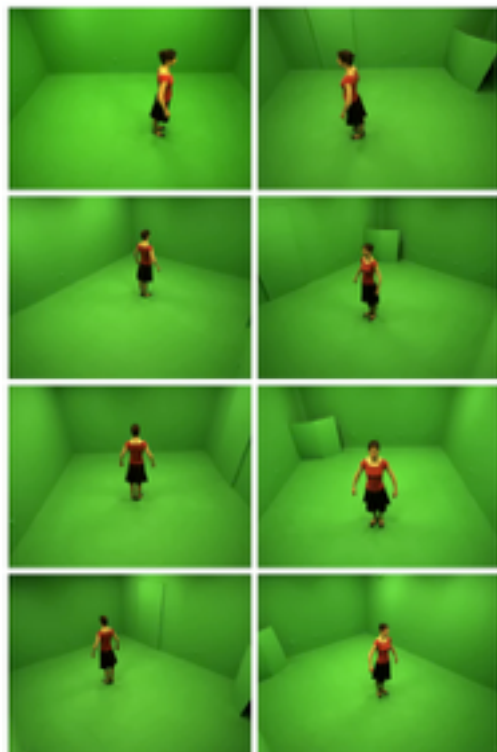
Vlasic 2008
Carranza 2003

Model-Free

- Flexible, no prior knowledge of scene required
- no specialist equipment only cameras

- Each frame is reconstructed independently, requires temporal alignment
- Model detail is limited by camera resolution

Starck 2007



Vlasic 2008



Model-Based - Vlasic 2008



Model-Free - Stark and Hilton 2007

Texturing 4D Models

Multi-View Video



~700MB




- ✓ View Dependent
- ✓ Dynamic
- ✗ Practical Storage

Texturing 4D Models

Multi-View Video



~700MB

-  View Dependent
-  Dynamic
-  Practical Storage

Texturing 4D Models

Multi-View Video



~700MB

- ✓ View Dependent
- ✓ Dynamic
- ✗ Practical Storage

Texturing 4D Models

Multi-View Video



~700MB

- ✓ View Dependent
- ✓ Dynamic
- ✗ Practical Storage

Single-Texture Map



~1MB

- ✗ View Dependent
- ✗ Dynamic
- ✓ Practical Storage

Texturing 4D Models

Multi-View Video



~700MB

- ✓ View Dependent
- ✓ Dynamic
- ✗ Practical Storage

Single-Texture Map



~1MB

- ✗ View Dependent
- ✗ Dynamic
- ✓ Practical Storage

Texturing 4D Models

Multi-View Video



~700MB

- ✓ View Dependent
- ✓ Dynamic
- ✗ Practical Storage

Single-Texture Map



~1MB

- ✗ View Dependent
- ✗ Dynamic
- ✓ Practical Storage

Texture map per Frame



~30MB

- ✗ View Dependent
- ✓ Dynamic
- ✓ Practical Storage

Texturing 4D Models

Multi-View Video



~700MB

- ✓ View Dependent
- ✓ Dynamic
- ✗ Practical Storage

Single-Texture Map



~1MB

- ✗ View Dependent
- ✗ Dynamic
- ✓ Practical Storage

Texture map per Frame

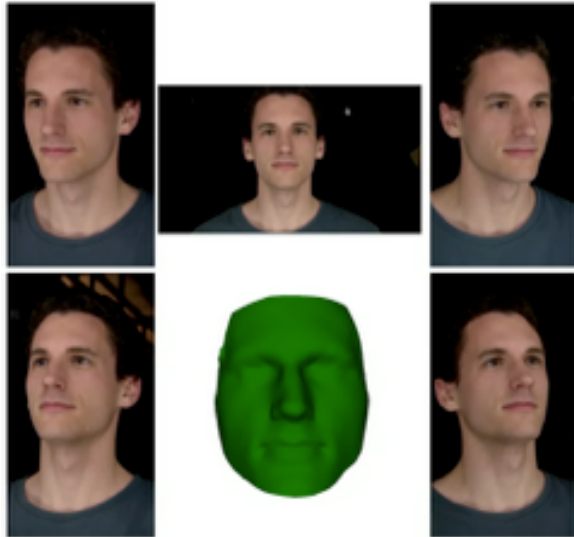


~30MB

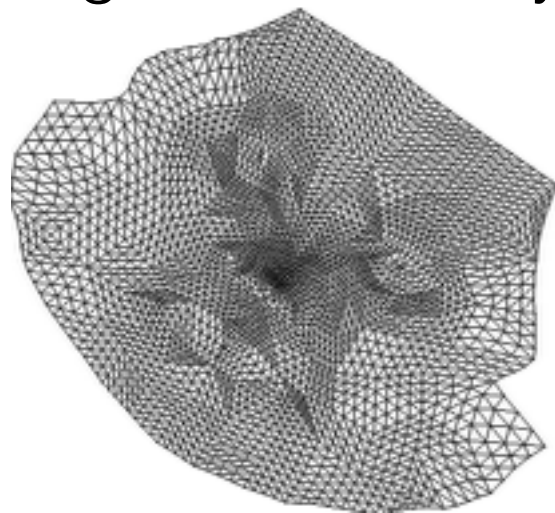
- ✗ View Dependent
- ✓ Dynamic
- ✓ Practical Storage

Multi-Layer Texture Map

INPUT



Multi-View Video
Aligned Geometry



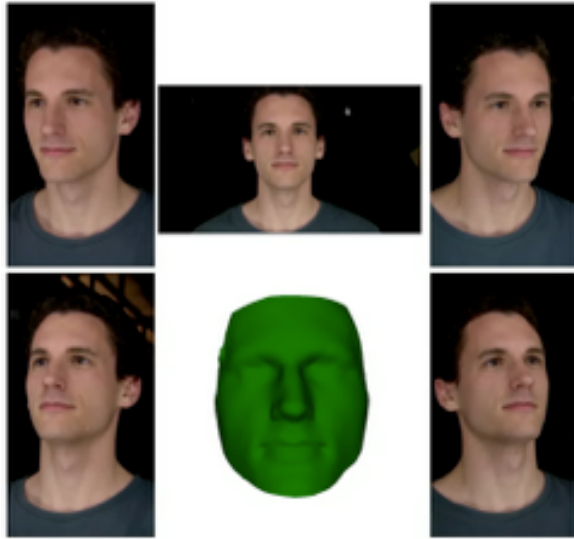
UV Coordinates

PROCESS

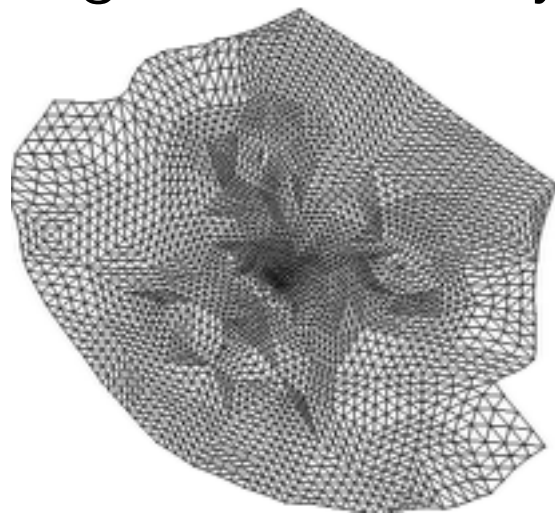
ASSIGNMENT

Multi-Layer Texture Map

INPUT



Multi-View Video
Aligned Geometry



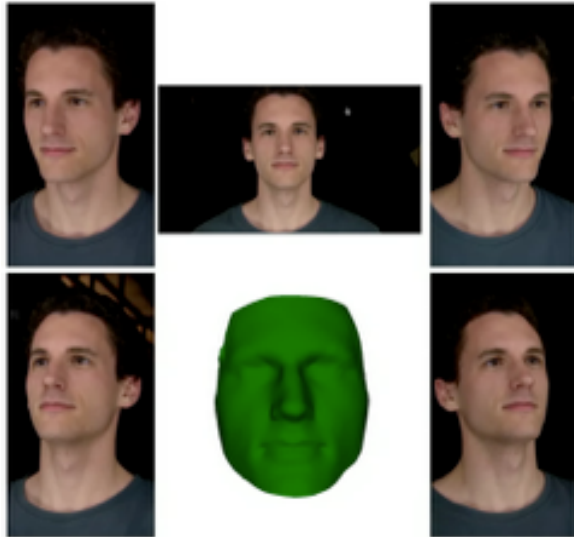
UV Coordinates

PROCESS

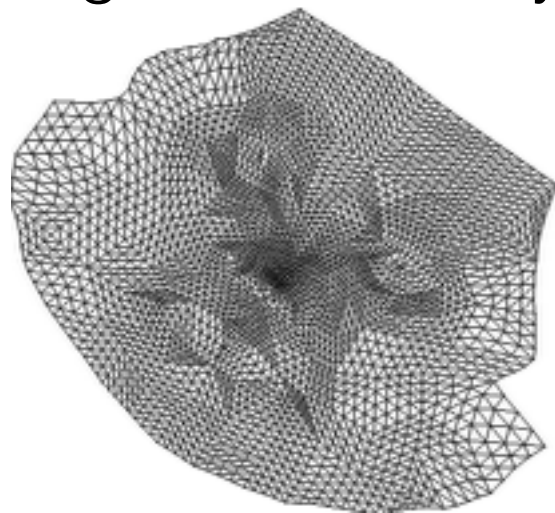
ASSIGNMENT

Multi-Layer Texture Map

INPUT

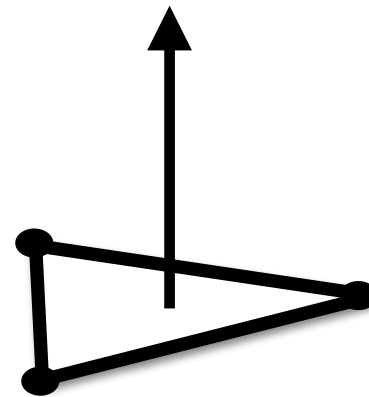


Multi-View Video
Aligned Geometry



UV Coordinates

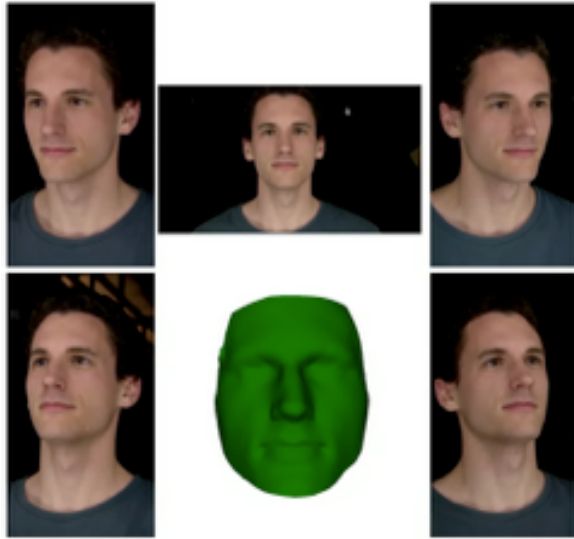
PROCESS



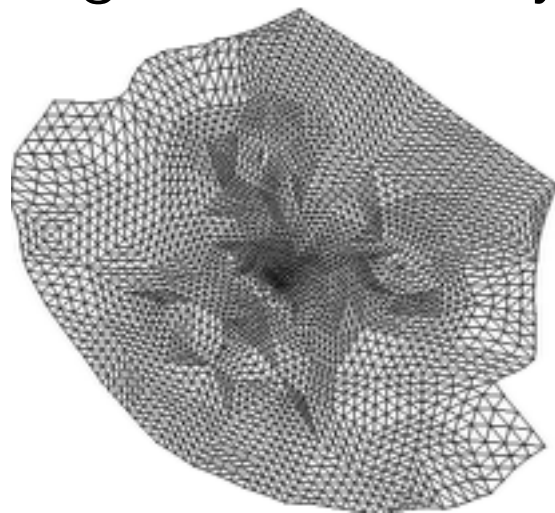
ASSIGNMENT

Multi-Layer Texture Map

INPUT

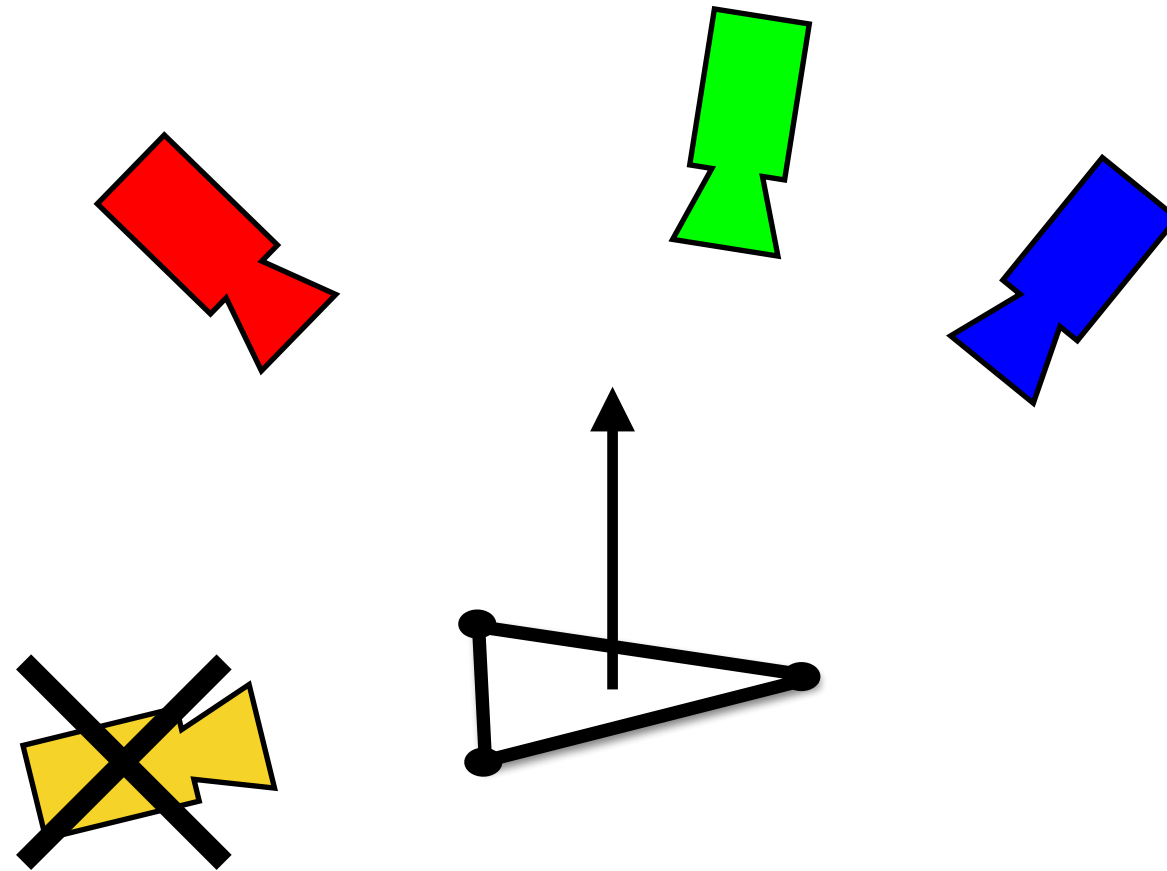


Multi-View Video
Aligned Geometry



UV Coordinates

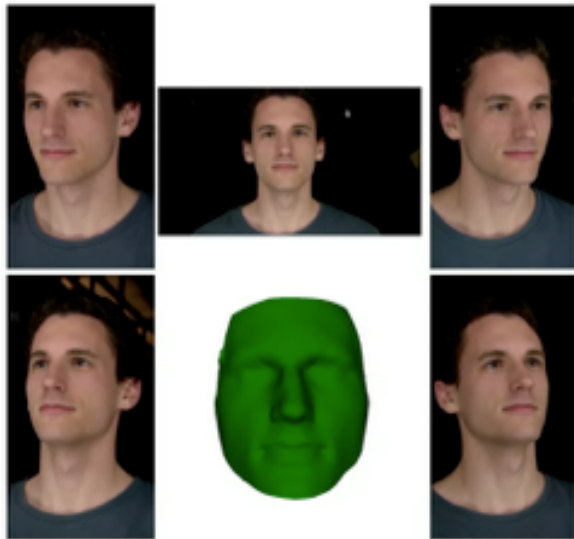
PROCESS



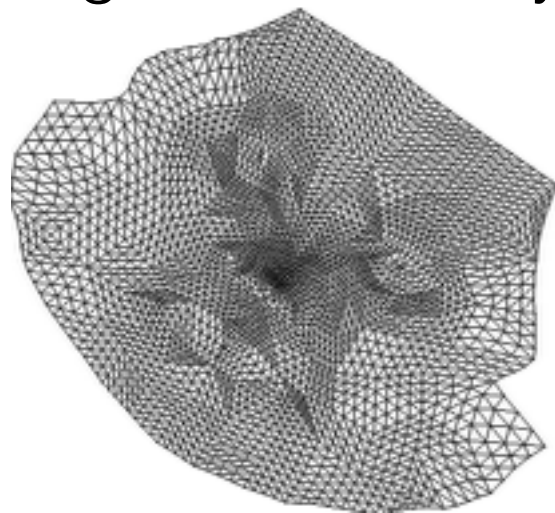
ASSIGNMENT

Multi-Layer Texture Map

INPUT

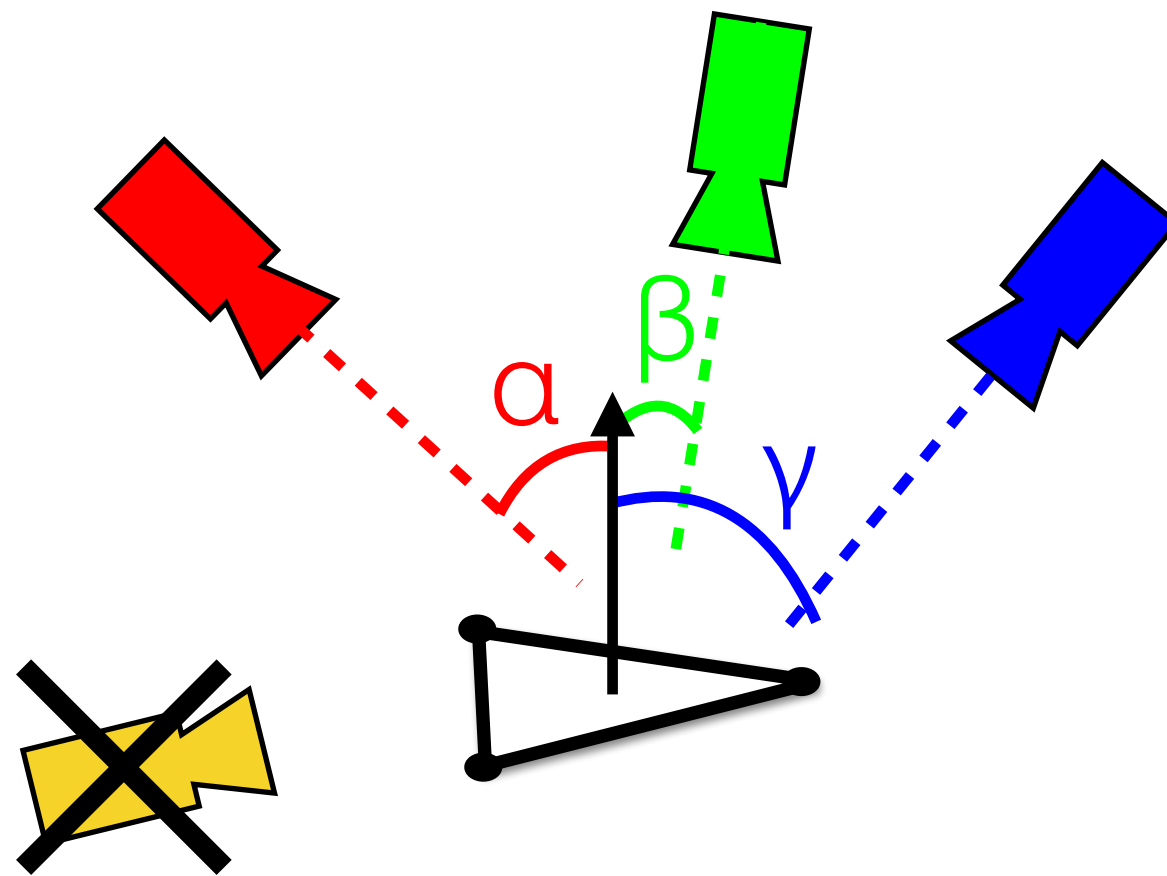


Multi-View Video
Aligned Geometry



UV Coordinates

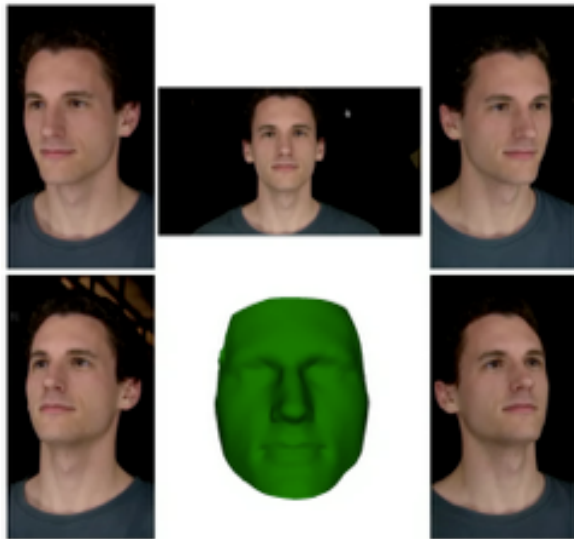
PROCESS



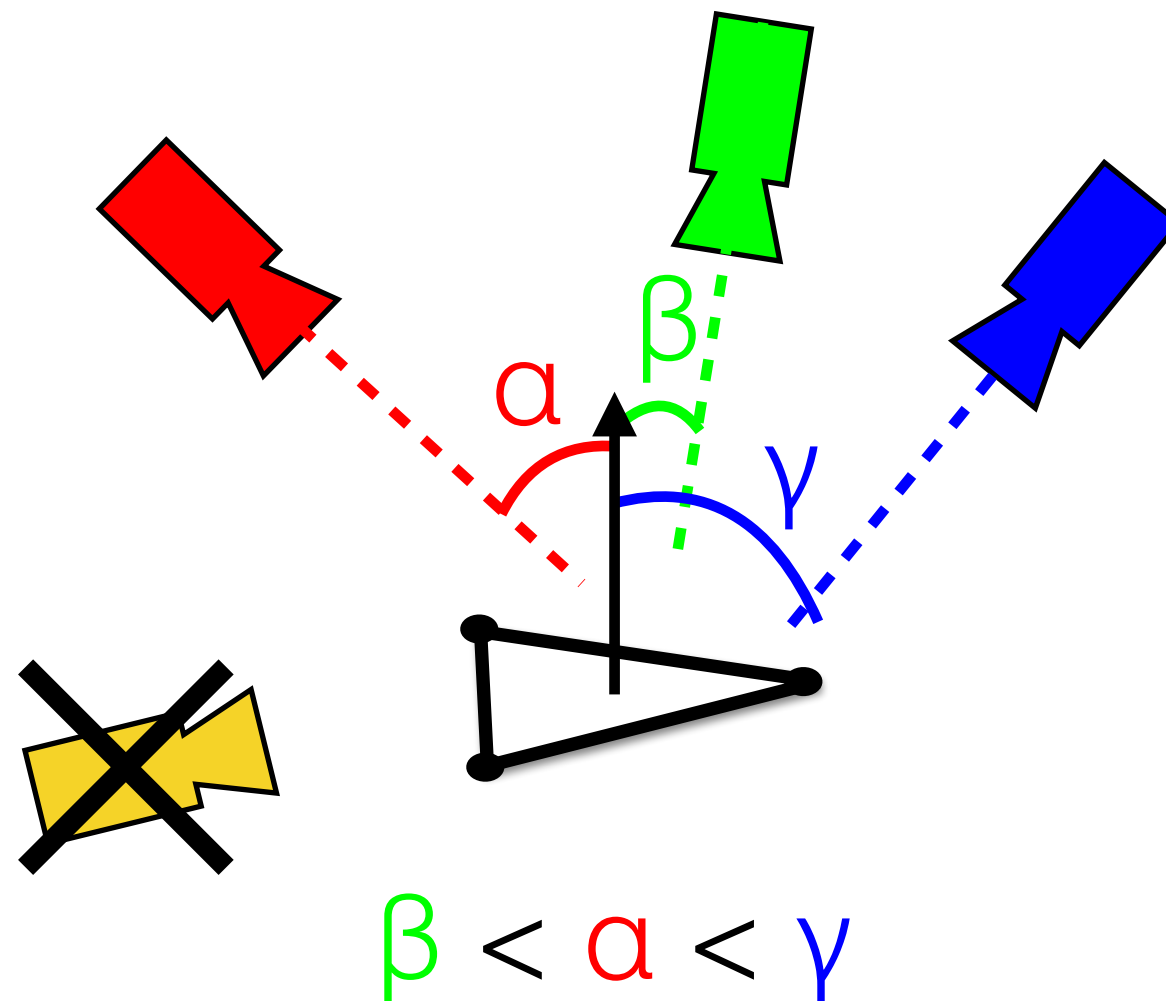
ASSIGNMENT

Multi-Layer Texture Map

INPUT



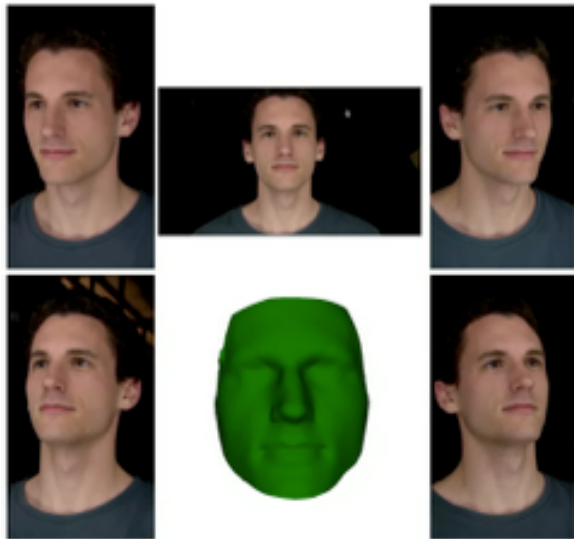
PROCESS



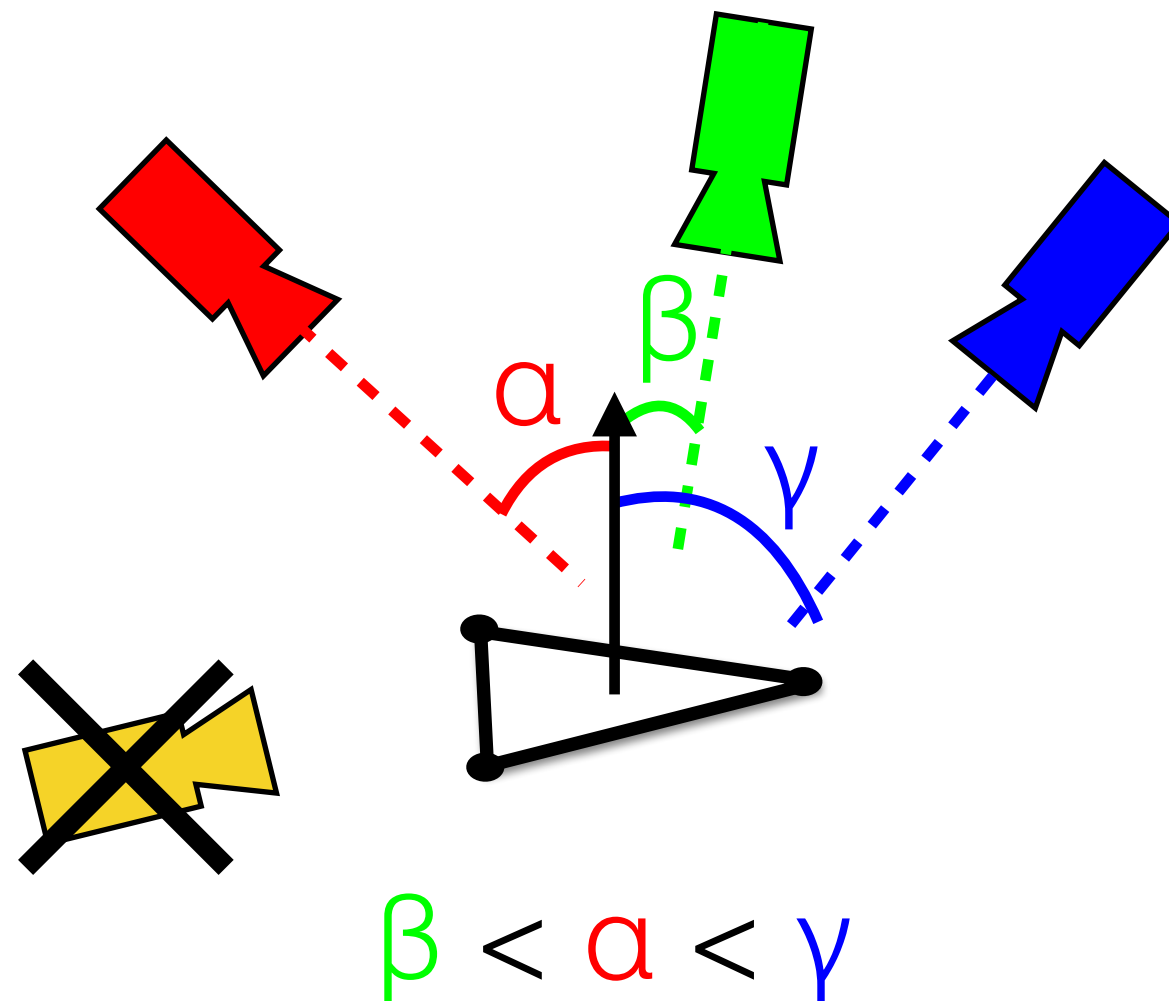
ASSIGNMENT

Multi-Layer Texture Map

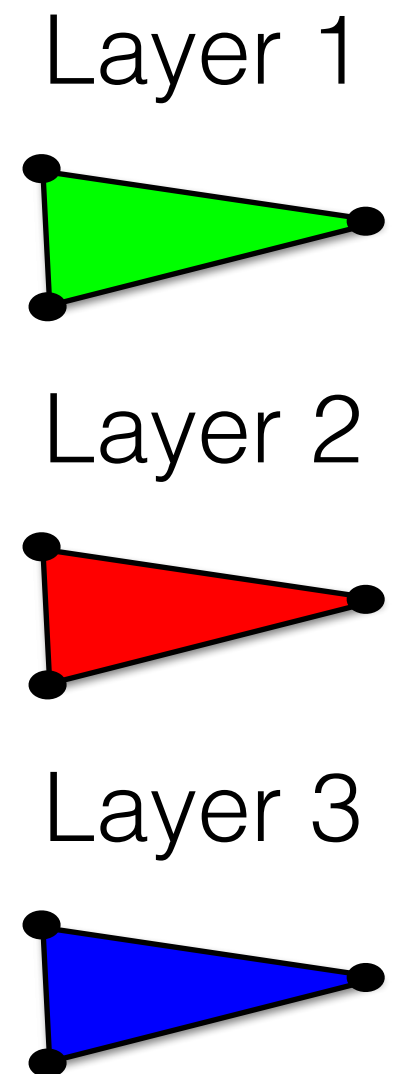
INPUT



PROCESS



ASSIGNMENT



Multi-Layer Texture Map

Layer 1

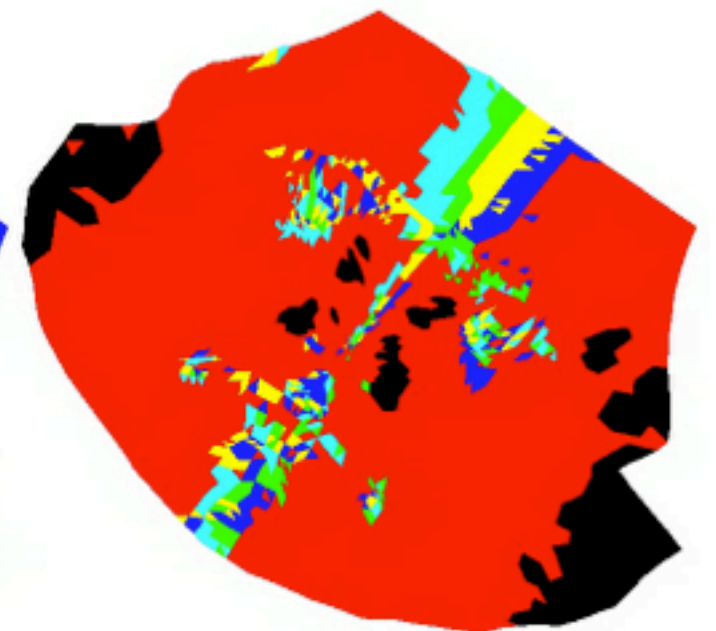
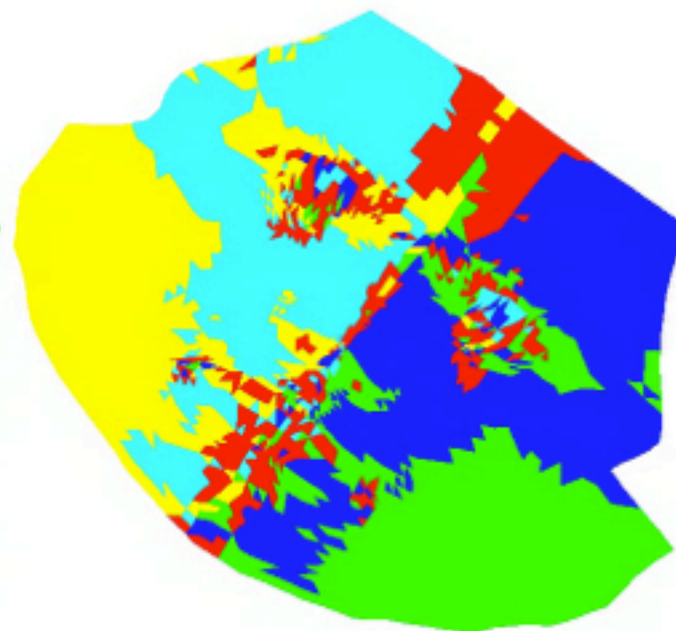
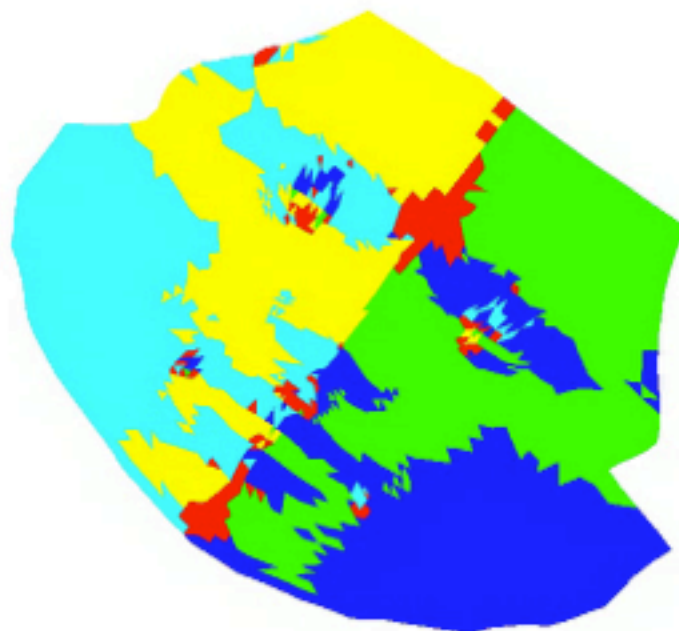
Layer 2

Layer 3

Texture
Colour



Camera
Assignment



Multi-Layer Texture Map

Layer 1

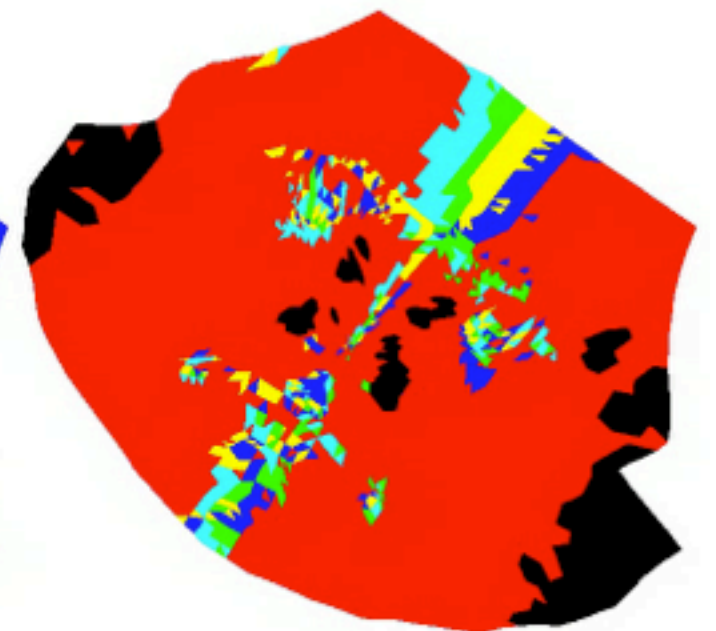
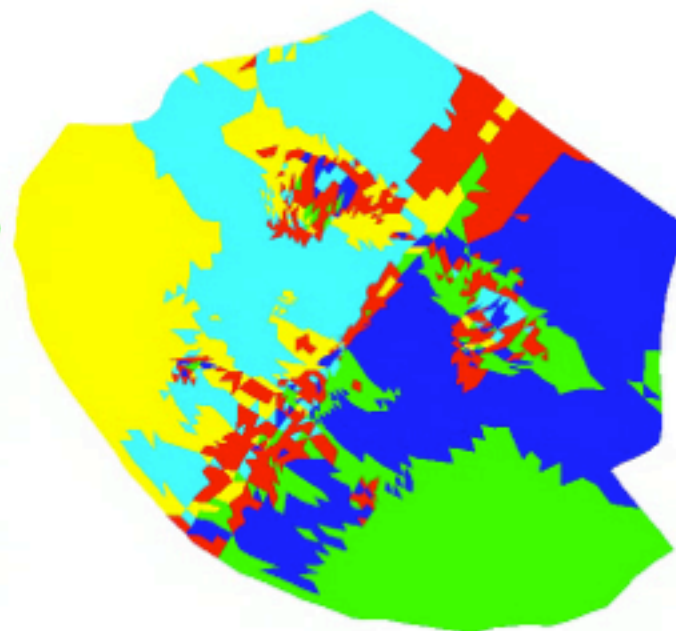
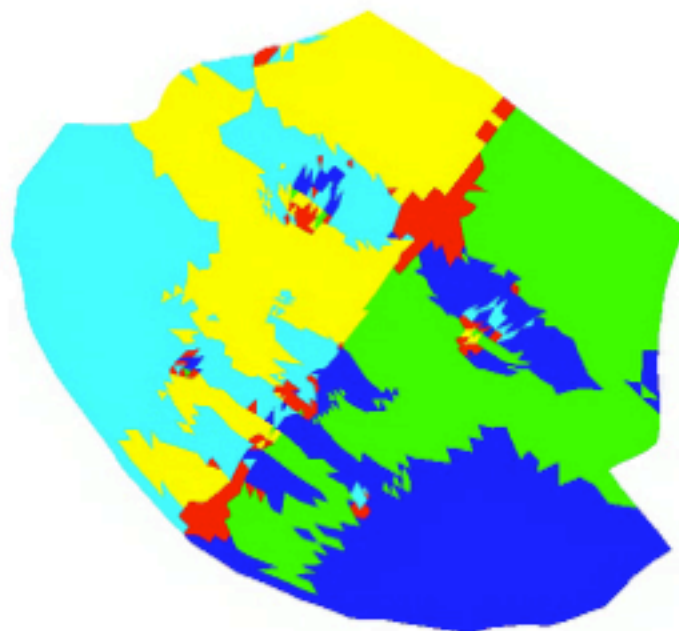
Layer 2

Layer 3

Texture
Colour

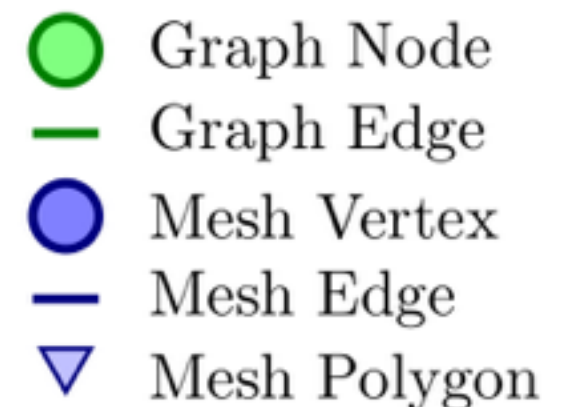
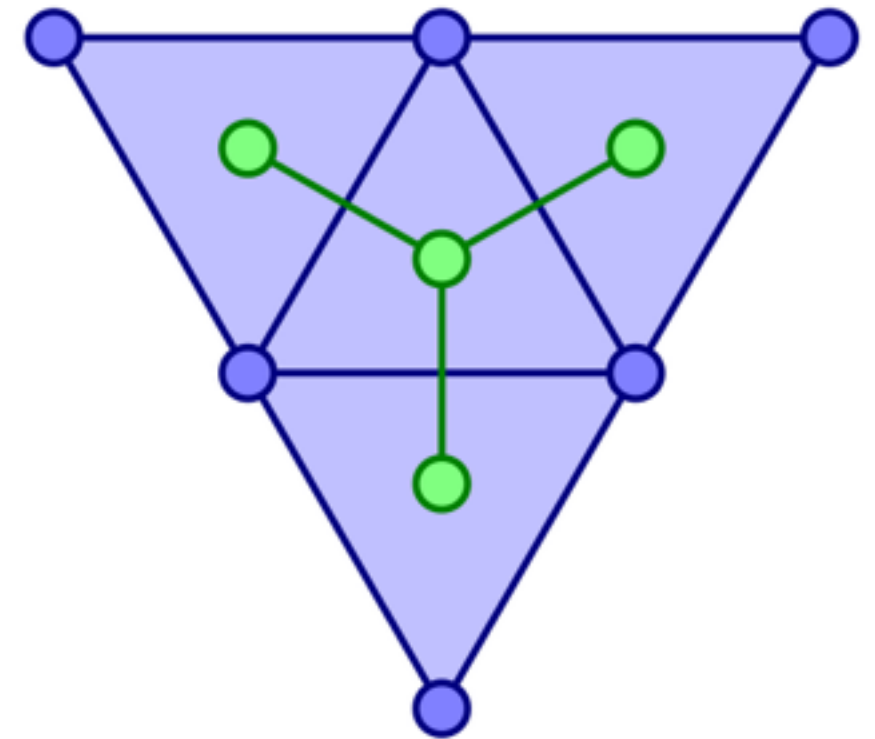


Camera
Assignment



Optimisation

- Construct an undirected graph
Nodes -> Mesh Polygons
Edges -> One Neighbour connection
- Using Markov Random Fields
find a labelling of cameras to
mesh polygons that minimises
an energy function



Optimisation

$$E(L(t)) = \sum_{\forall t} (E_v(L(t)) + \lambda_s E_s(L(t)) + \lambda_t E_t(L(t), L(t+1)))$$

Unary Term

$$E(L(t)) = \sum_{\forall t} \boxed{E_v(L(t))} + \lambda_s E_s(L(t)) + \lambda_t E_t(L(t), L(t+1))$$

- Unary Term
Ensures most direct camera is preferred and enforces visibility constraints

$$E(L(t)) = \sum_{\forall t} \left(\overset{\text{Unary Term}}{E_v(L(t))} + \lambda_s \overset{\text{Spatial Term}}{\boxed{E_s(L(t))}} + \lambda_t E_t(L(t), L(t+1)) \right)$$

- Unary Term
Ensures most direct camera is preferred and enforces visibility constraints
- Spatial Smoothness Term
Reduces changes in camera-to-polygon assignment across mesh surface

Optimisation

$$E(L(t)) = \sum_{\forall t} (\overset{\text{Unary Term}}{E_v(L(t))} + \overset{\text{Spatial Term}}{\lambda_s E_s(L(t))} + \overset{\text{Temporal Term}}{\lambda_t E_t(L(t), L(t+1))})$$

- Unary Term
Ensures most direct camera is preferred and enforces visibility constraints
- Spatial Smoothness Term
Reduces changes in camera-to-polygon assignment across mesh surface
- Temporal Smoothness Term
Reduces changes in camera-to-polygon assignment over time

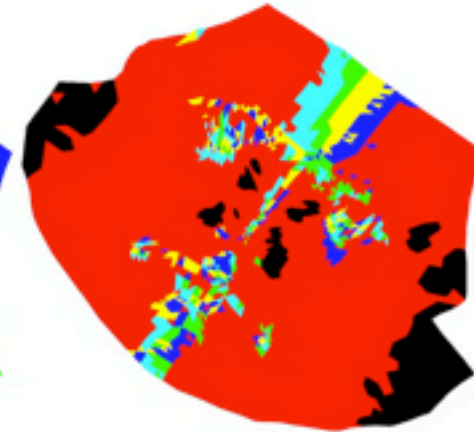
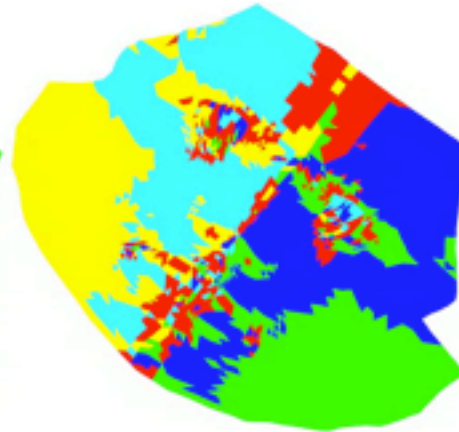
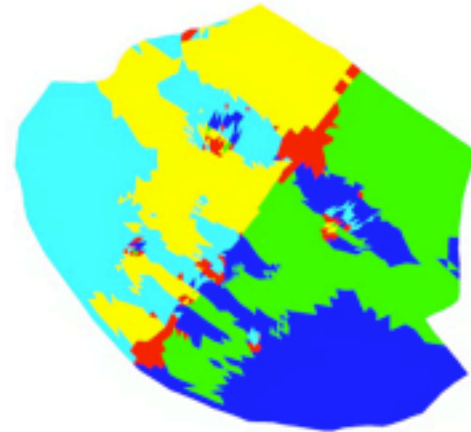
Optimisation

Layer 1

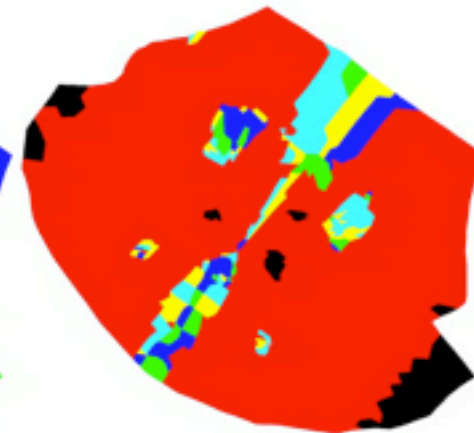
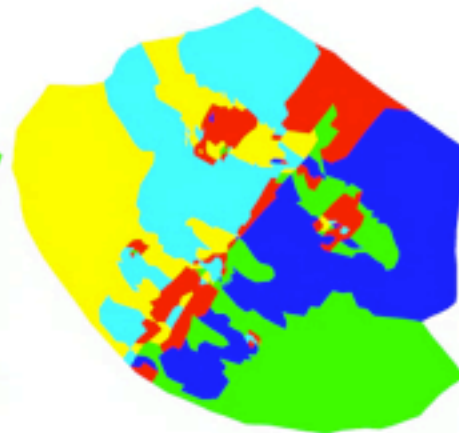
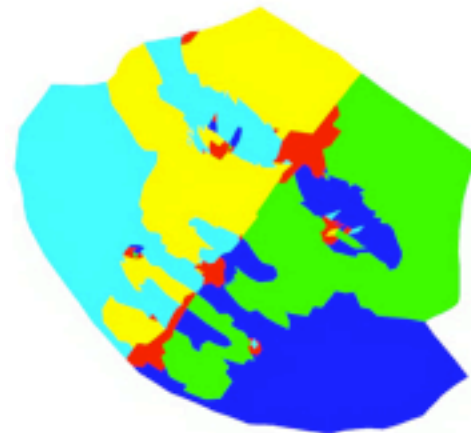
Layer 2

Layer 3

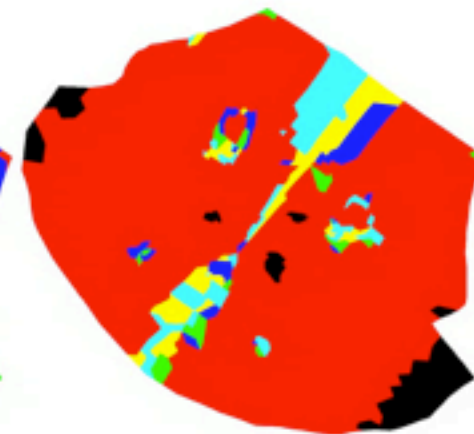
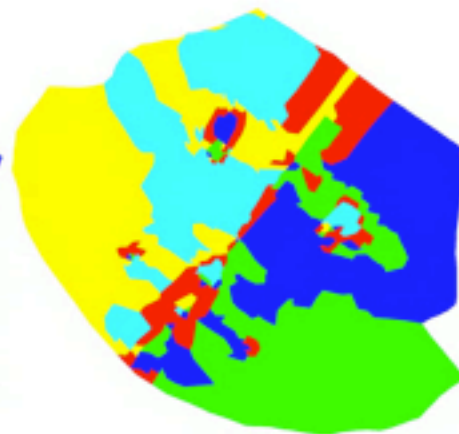
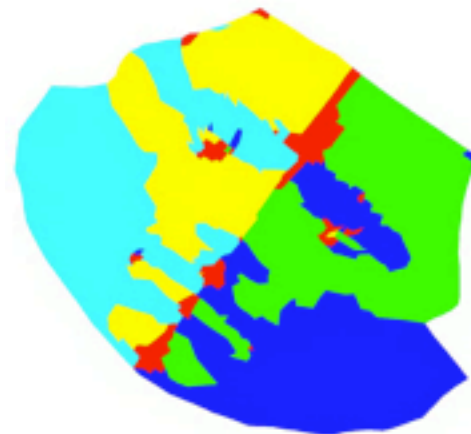
No
Optimisation
NO



Spatial
Optimisation
SO



Spatio + Temporal
Optimisation
TO



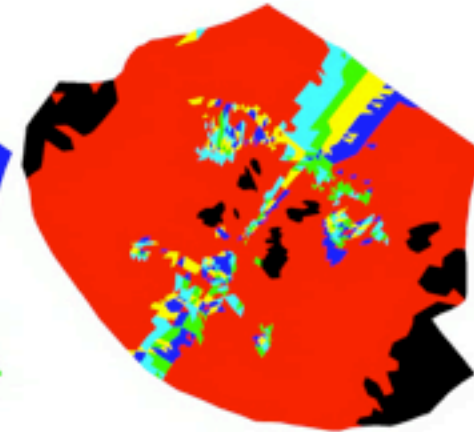
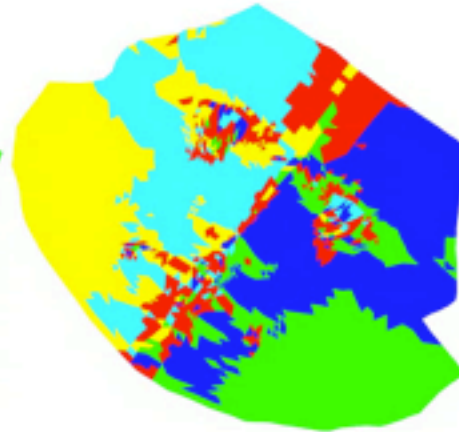
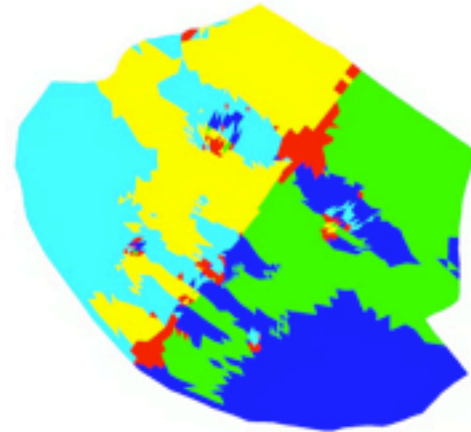
Optimisation

Layer 1

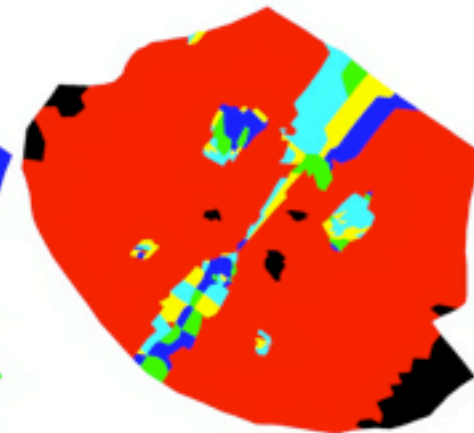
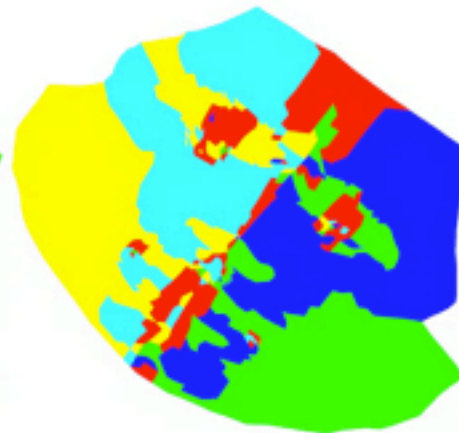
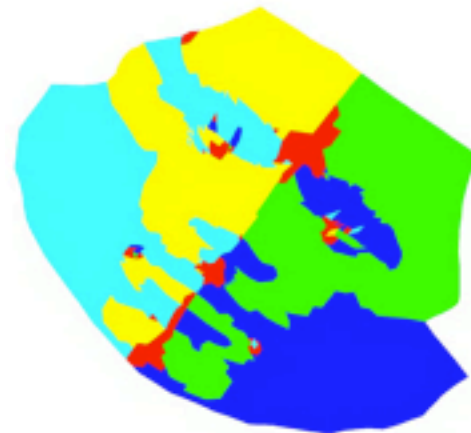
Layer 2

Layer 3

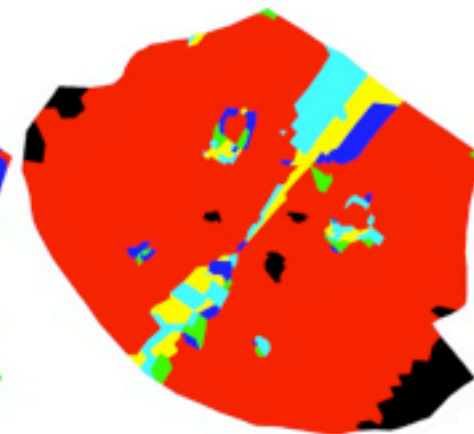
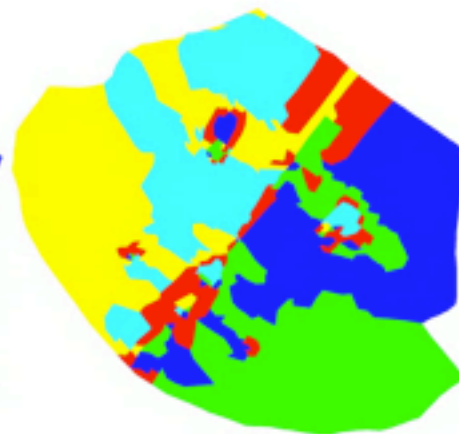
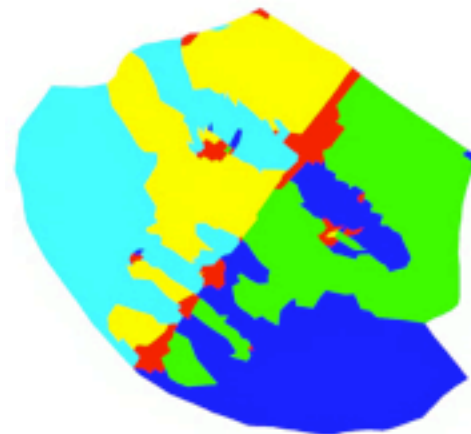
No
Optimisation
NO



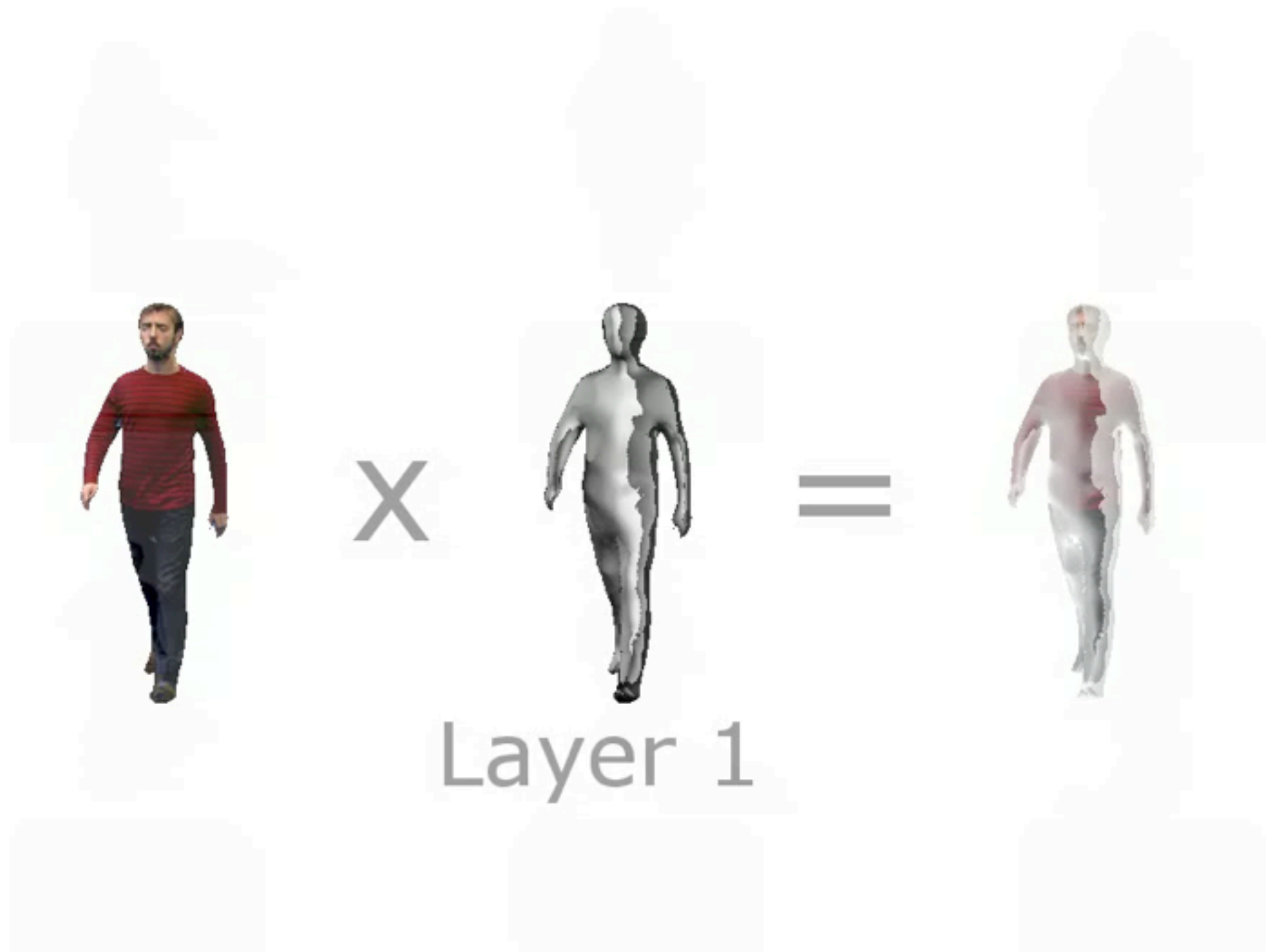
Spatial
Optimisation
SO



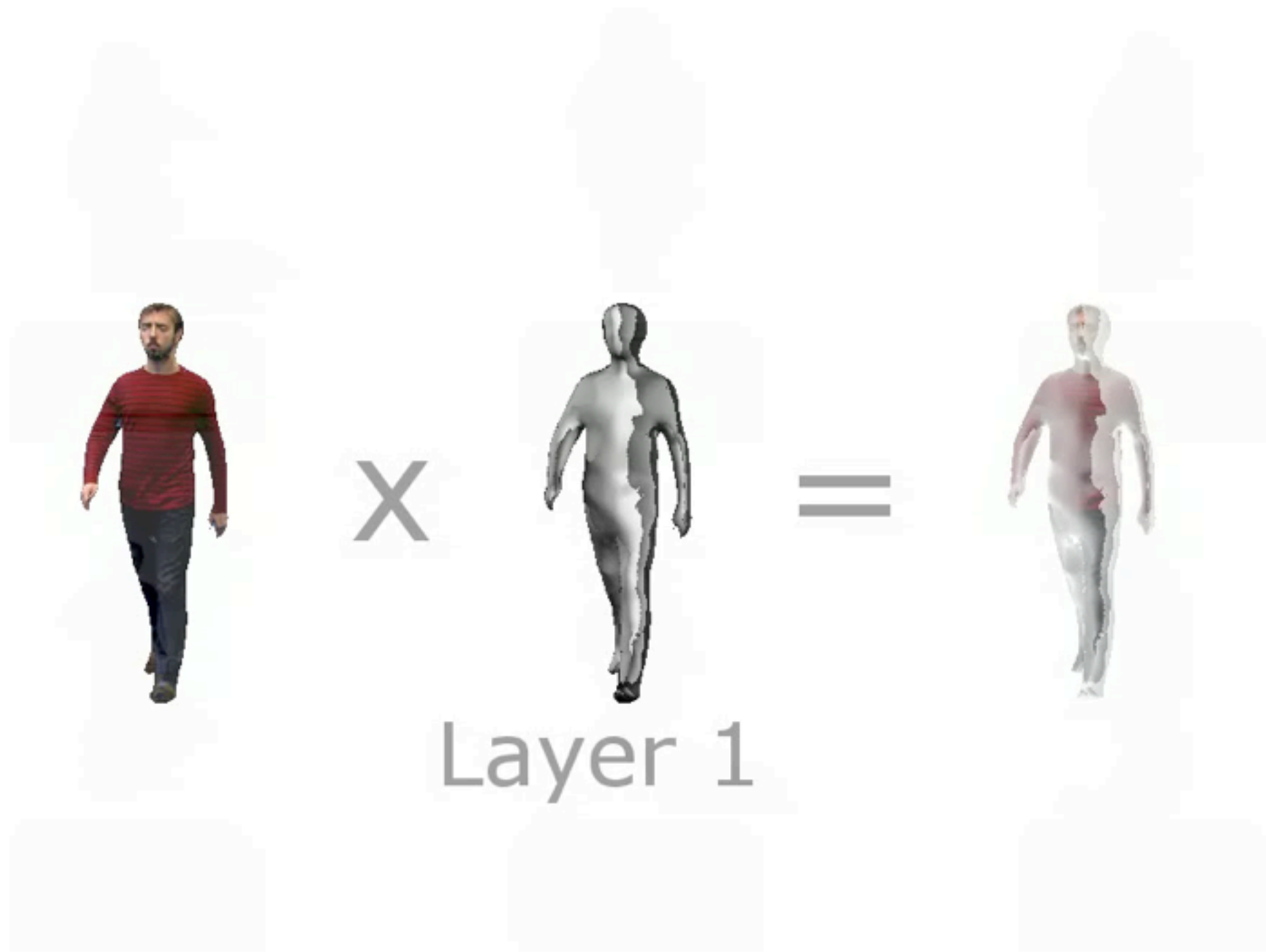
Spatio + Temporal
Optimisation
TO



Rendering

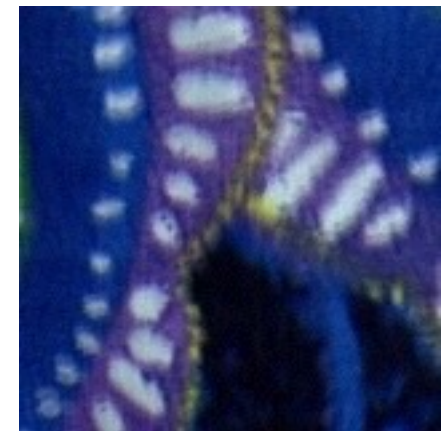
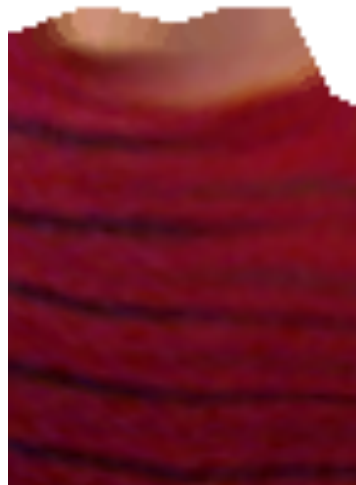


Rendering



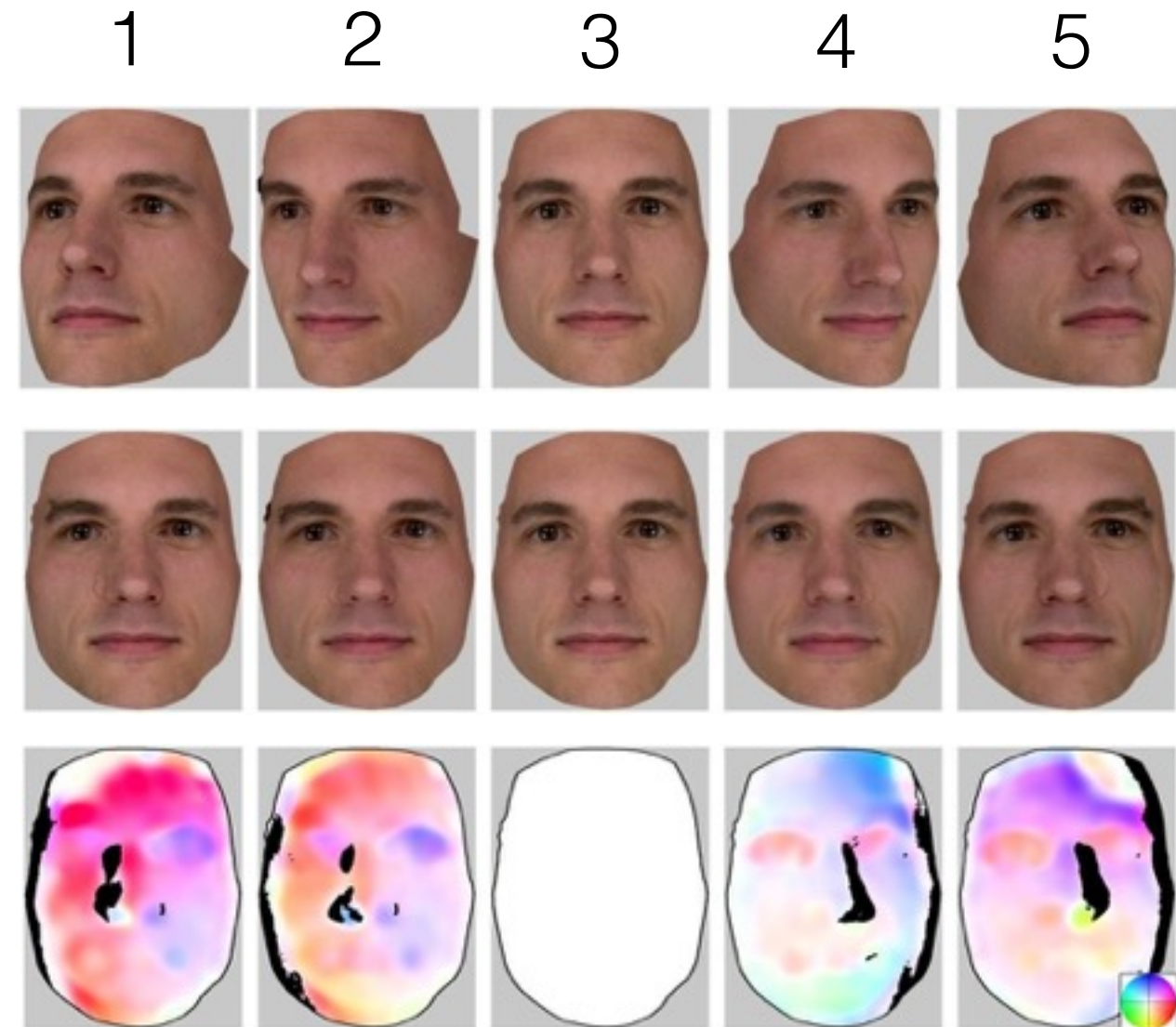
Texture Artefacts

- Approximate geometry and imprecise camera calibration leads to artefacts when blending textures as in view-dependent rendering
- Highlights need for spatial alignment



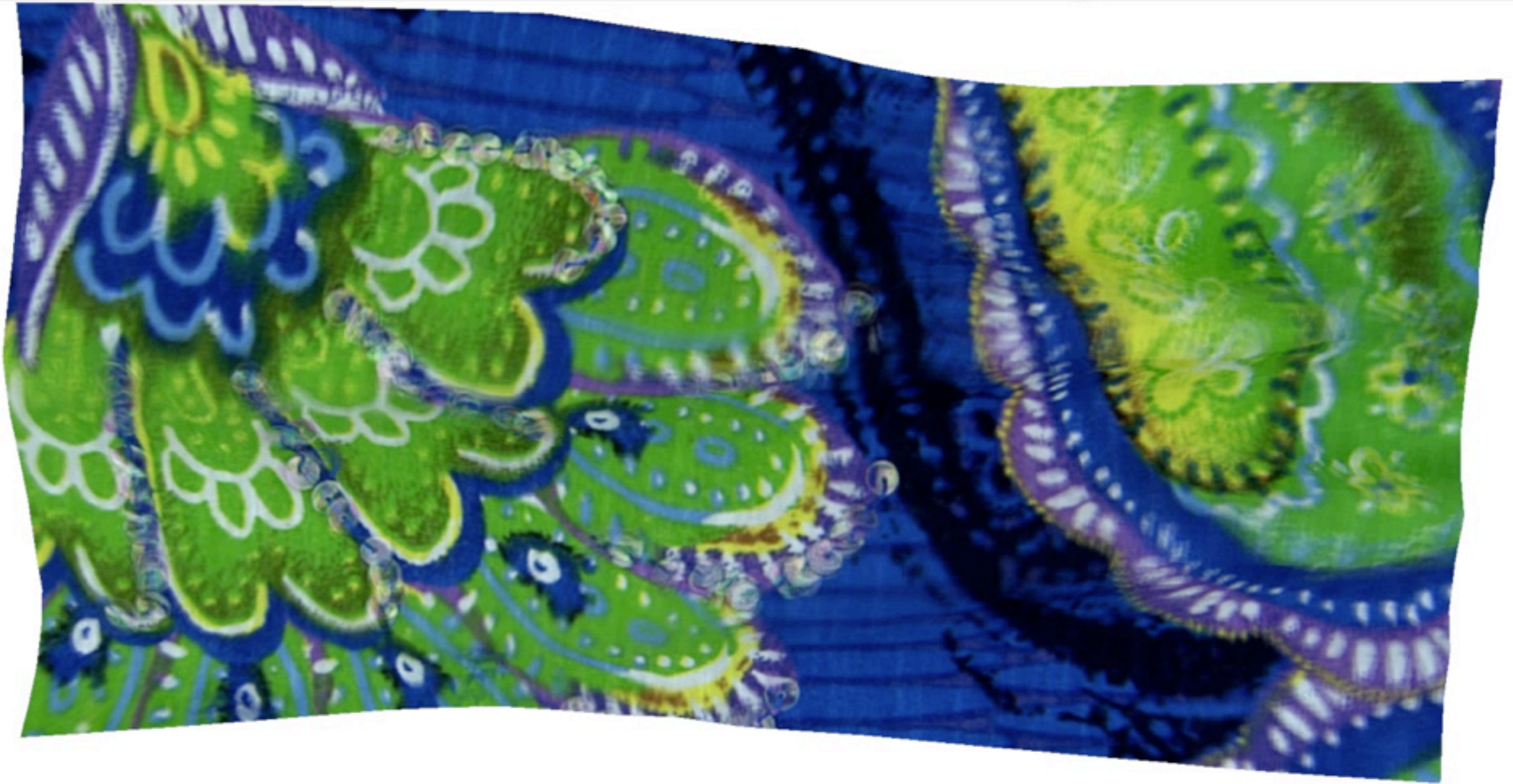
Multi-View Texture Alignment

- From viewpoint of each camera, projectively texture from all other cameras and compute optical flow
- Optical flow breaks down in presence of occlusions
- Discard flow in areas of occlusions and depth discontinuities (black areas)

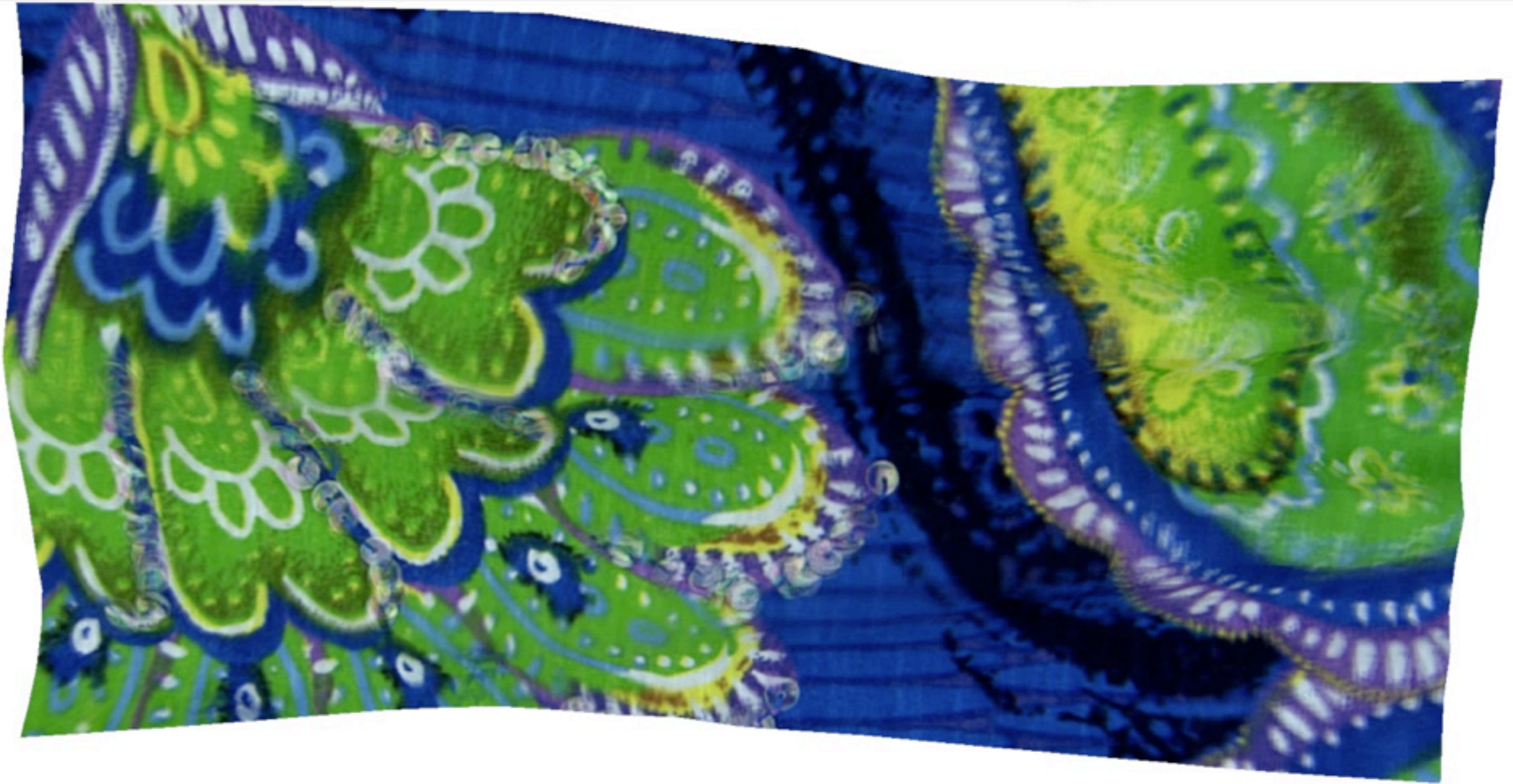


$$V_i = \sum_{j=1}^{N_C} \omega_j S_i^j O_{i \rightarrow j}$$

Without multi-view texture alignment



Without multi-view texture alignment



Without multi-view texture alignment

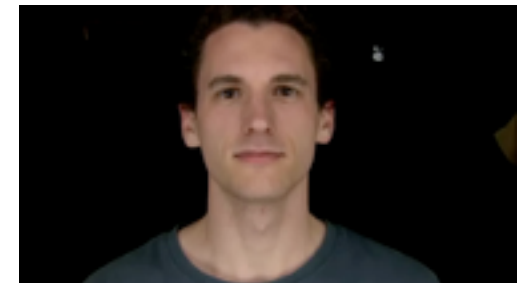


Without multi-view texture alignment



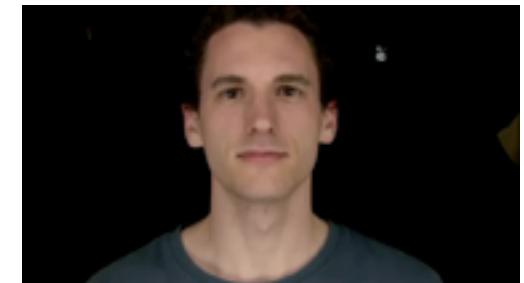
Evaluation: Datasets

Dataset	Cameras	Frames	Captured Data (MB)	
			Raw	Video Compression
Character 1	8	31	1800	61
Cloth	5	310	11400	906
Dan	8	28	1600	57
Face	5	355	13100	386



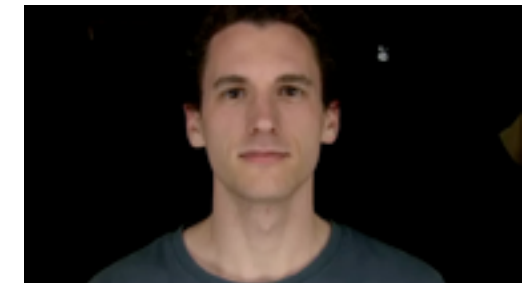
Evaluation: Datasets

Dataset	Cameras	Frames	Captured Data (MB)	
			Raw	Video Compression
Character 1	8	31	1800	61
Cloth	5	310	11400	906
Dan	8	28	1600	57
Face	5	355	13100	386



Evaluation: Datasets

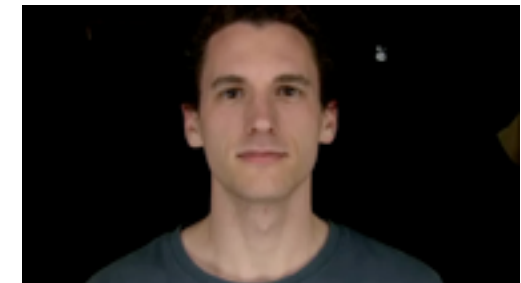
Dataset	Cameras	Frames	Captured Data (MB)	
			Raw	Video Compression
Character 1	8	31	1800	61
Cloth	5	310	11400	906
Dan	8	28	1600	57
Face	5	355	13100	386



- Compare storage of layer sequence to the video compressed captured data both encoded using same codec

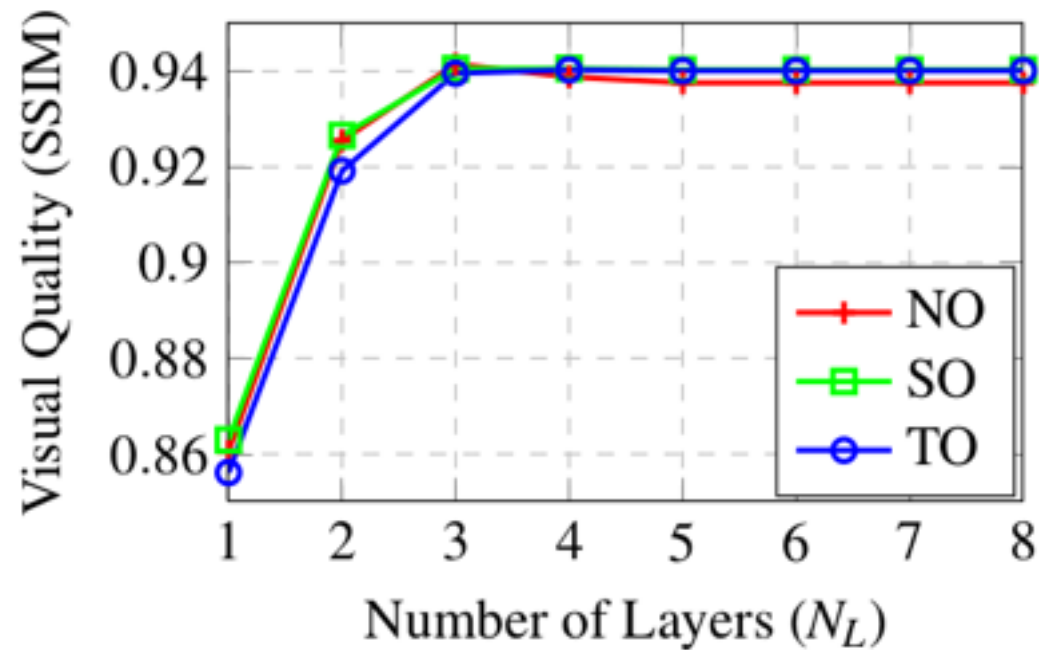
Evaluation: Datasets

Dataset	Cameras	Frames	Captured Data (MB)	
			Raw	Video Compression
Character 1	8	31	1800	61
Cloth	5	310	11400	906
Dan	8	28	1600	57
Face	5	355	13100	386



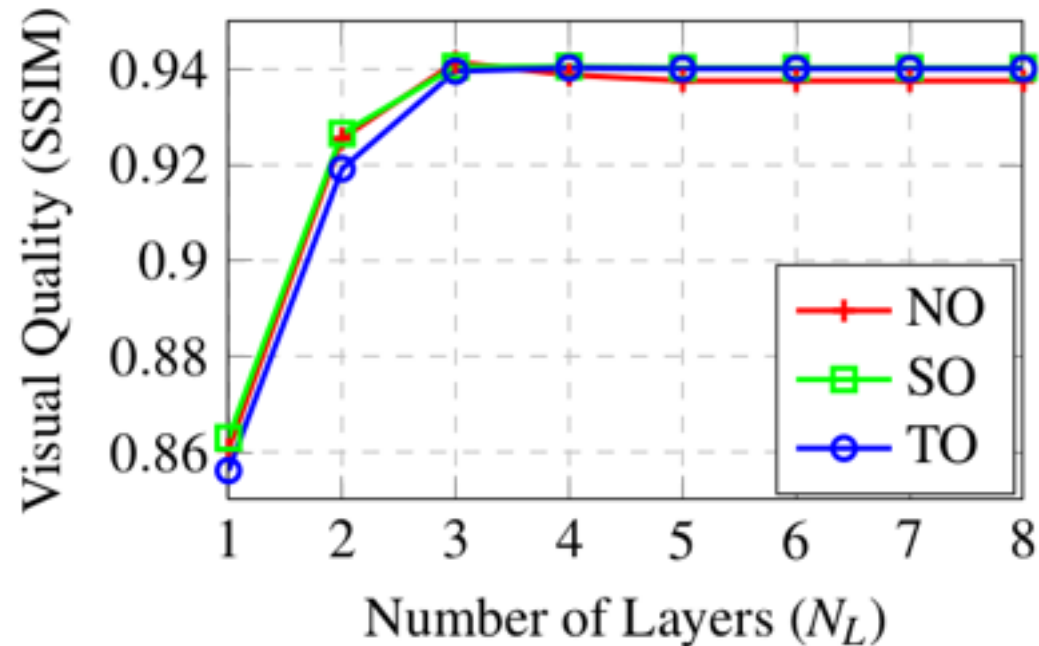
- Compare storage of layer sequence to the video compressed captured data both encoded using same codec
- Evaluate quality by rendering arbitrary views using multi-layer texture representation and using free viewpoint video renderer [Starck et al. 2009] and compare using Structural Similarity Index Measure [Wang et al. 2009].

Evaluation: Quality



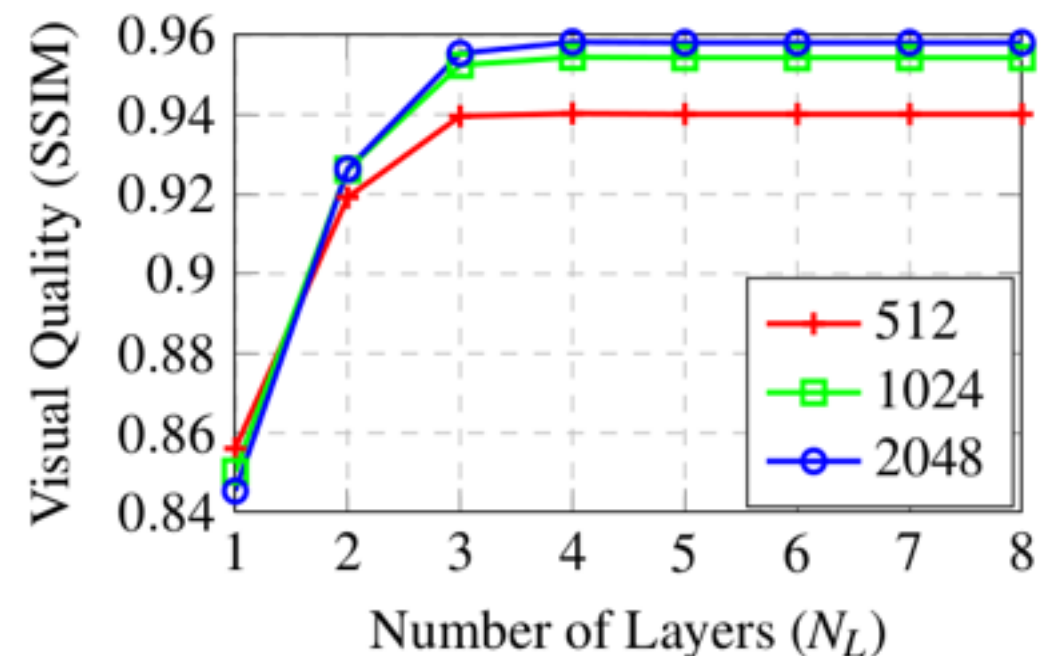
- Optimisation has no effect on rendering quality
- Same appearance information re-ordered for better compression
- Result generated using 512 texture size

Evaluation: Quality

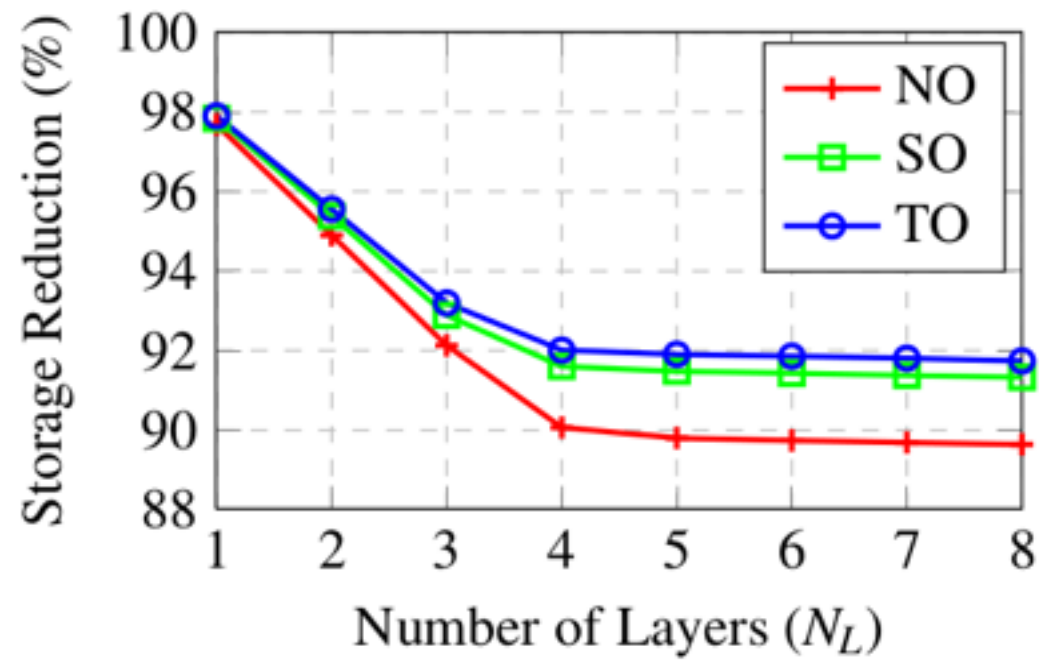


- Optimisation has no effect on rendering quality
- Same appearance information re-ordered for better compression
- Result generated using 512 texture size

- Increasing the size of the texture map has no effect on the quality above 1024
- Results generated using TO

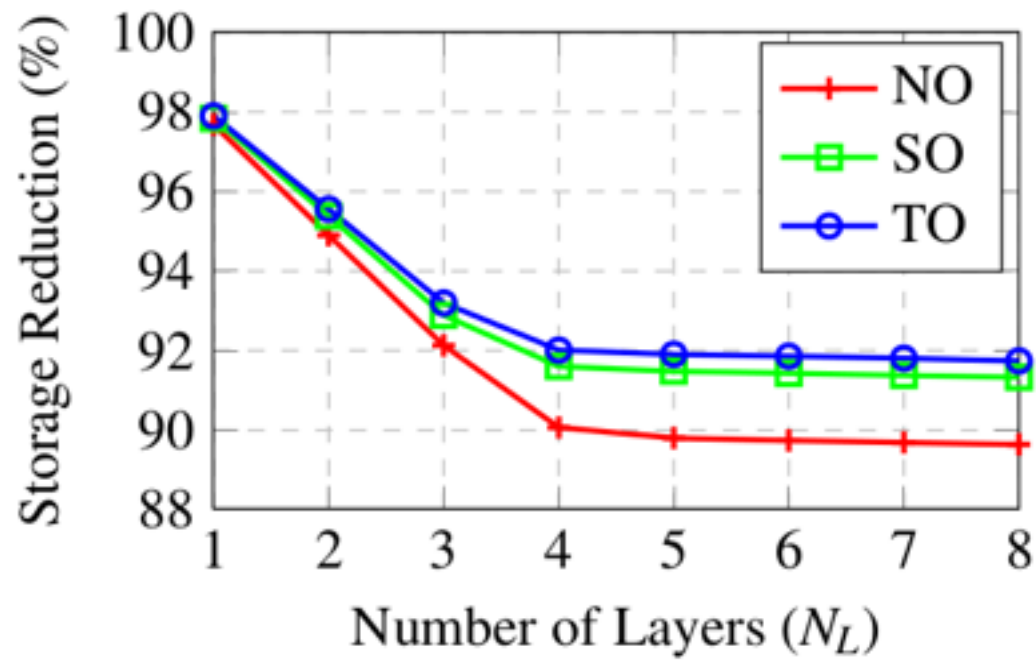


Evaluation: Storage



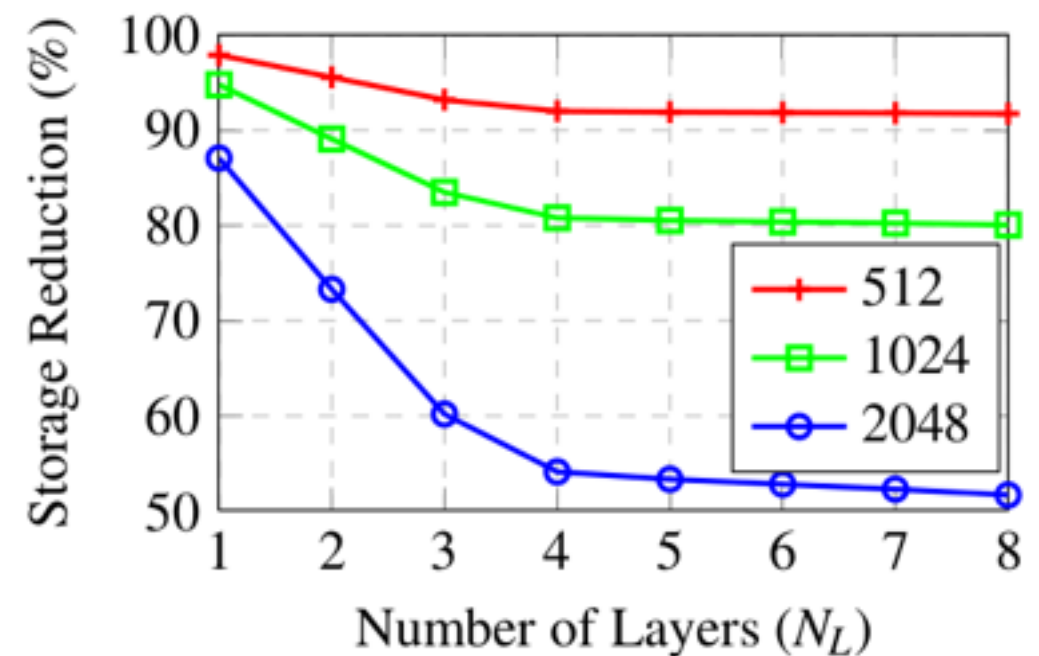
- Effect of optimisation on storage size
- With Face dataset this difference is approximately 10-20MB
- Results generated using 512 texture size

Evaluation: Storage



- Effect of optimisation on storage size
- With Face dataset this difference is approximately 10-20MB
- Results generated using 512 texture size

- Higher the texture size, the lower the storage reduction
- In the case of datasets dan and character 1 reduction stop after 4 layers
- Results generated using TO

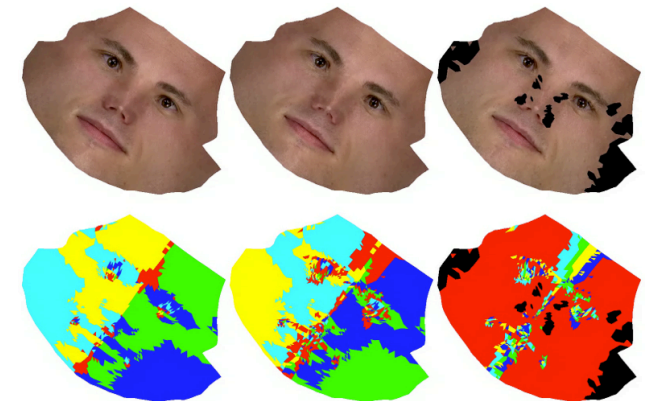


Conclusions

Conclusions

- Novel texture representation maintains dynamic, view-dependent appearance
- Demonstrated on a variety of subjects including full body, face and cloth capture

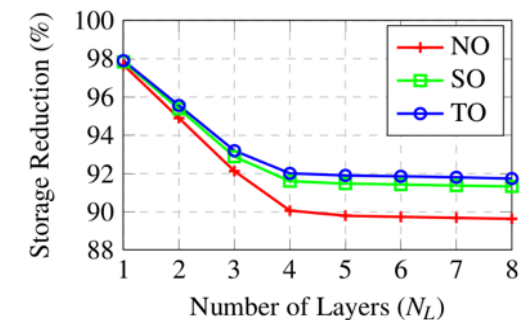
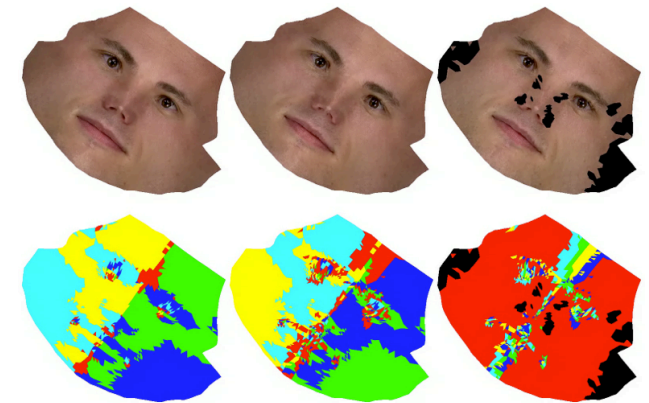
- ✓ View Dependent
- ✓ Dynamic
- ✓ Practical Storage



Conclusions

- Novel texture representation maintains dynamic, view-dependent appearance
- Demonstrated on a variety of subjects including full body, face and cloth capture
- Significant reduction in storage requirements $> 90\%$

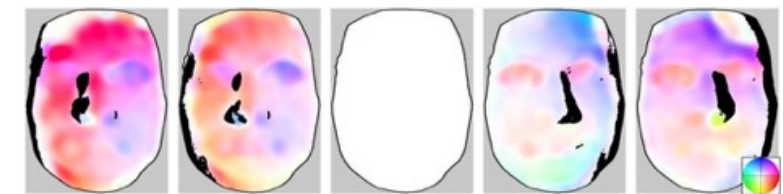
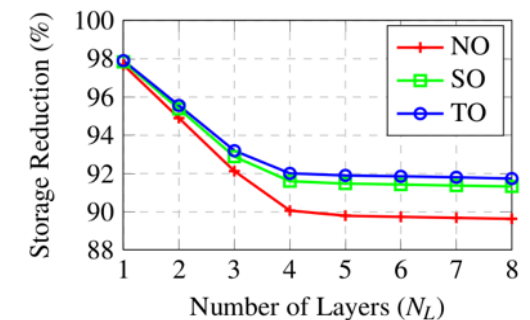
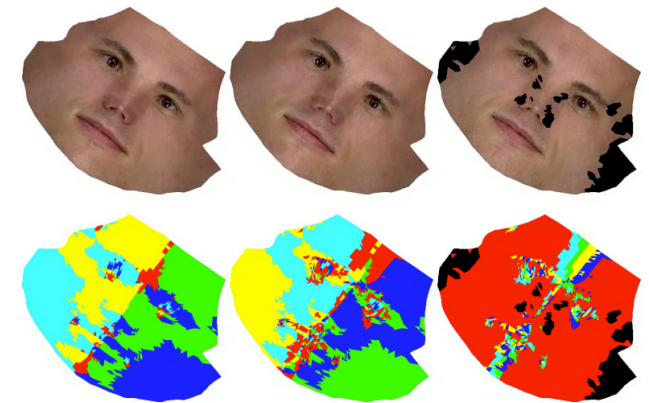
- ✓ View Dependent
- ✓ Dynamic
- ✓ Practical Storage



Conclusions

- Novel texture representation maintains dynamic, view-dependent appearance
- Demonstrated on a variety of subjects including full body, face and cloth capture
- Significant reduction in storage requirements $> 90\%$
- Optical flow based multi camera alignment which significantly reduces artefacts

- ✓ View Dependent
- ✓ Dynamic
- ✓ Practical Storage



Future Work

- Can we ensure temporal coherence over layer sequences?
- Can the multi-layer texture representation be extended to allow appearance editing and relighting through extraction of material properties?

CVSSP3D Data Directory

[Home](#) [Publications](#) [Datasets](#) [Register](#) [Contact](#)

Dataset: Dan

Recorded as part of Dans PhD



Box Low Big

8 Cameras, 162 Frames

Download: [Calibration Images](#)
[Silhouette](#)



Box Low Small

8 Cameras, 171 Frames

Download: [Calibration Images](#)
[Silhouette](#)

cvssp.org/data/cvssp3d

[Silhouette](#)

[Silhouette](#)



Jump Long

8 Cameras, 53 Frames

Download: [Calibration Images](#)
[Silhouette](#)



Jump Low

8 Cameras, 18 Frames

Download: [Calibration Images](#)
[Silhouette](#)



Jump Short

8 Cameras, 36 Frames

Download: [Calibration Images](#)
[Silhouette](#)



Turn Left

8 Cameras, 27 Frames

Download: [Calibration Images](#)
[Silhouette](#)



Turn Right

8 Cameras, 27 Frames

Download: [Calibration Images](#)
[Silhouette](#)



Walk

8 Cameras, 29 Frames

Download: [Calibration Images](#)
[Silhouette](#)

Acknowledgements

University of Surrey



Adrian Hilton

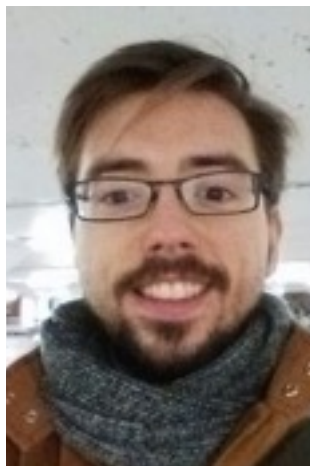


John Collomosse

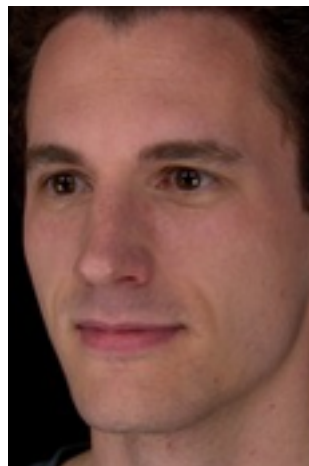
BBC R&D



Graham Thomas



Dan Casas*



Martin Kloudiny



Peng Huang

Project



* Now at USC: ICT

cvssp.org/data/cvssp3d

Thanks for your attention!
Questions ?



TO, 1024, 3 Layers
72 MB, 82% Reduction



TO, 512, 3 Layers
3.8 MB, 93% Reduction

cvssp.org/data/cvssp3d

Thanks for your attention!
Questions ?



TO, 1024, 3 Layers
72 MB, 82% Reduction



TO, 512, 3 Layers
3.8 MB, 93% Reduction