IEEE International Symposium on Multimedia 2012

A Variational Bayesian Inference Framework for Multiview Depth Image Enhancement

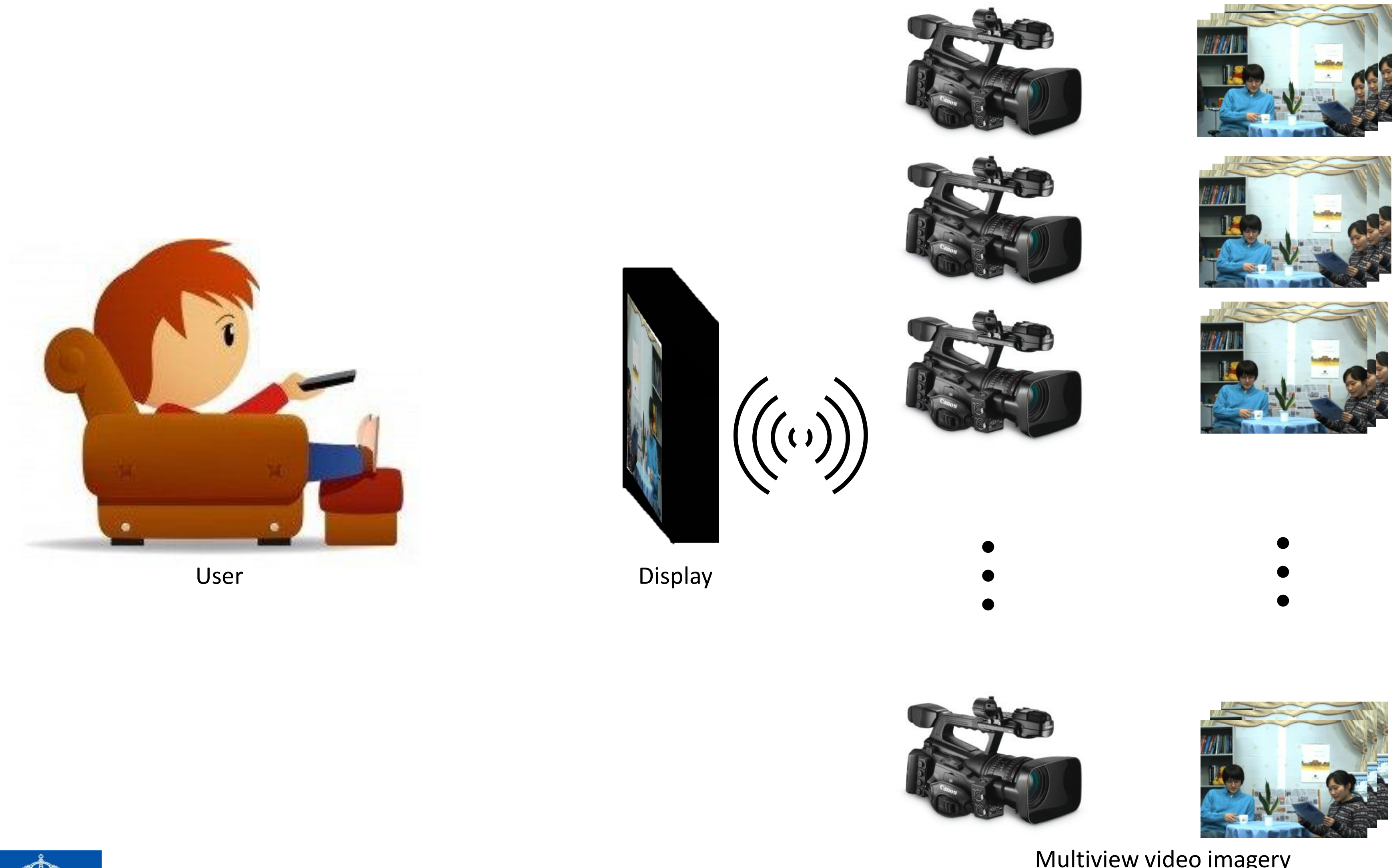
Pravin Kumar Rana, Jalil Taghia, and Markus Flierl

School of Electrical Engineering
KTH Royal Institute of Technology
Stockholm, Sweden

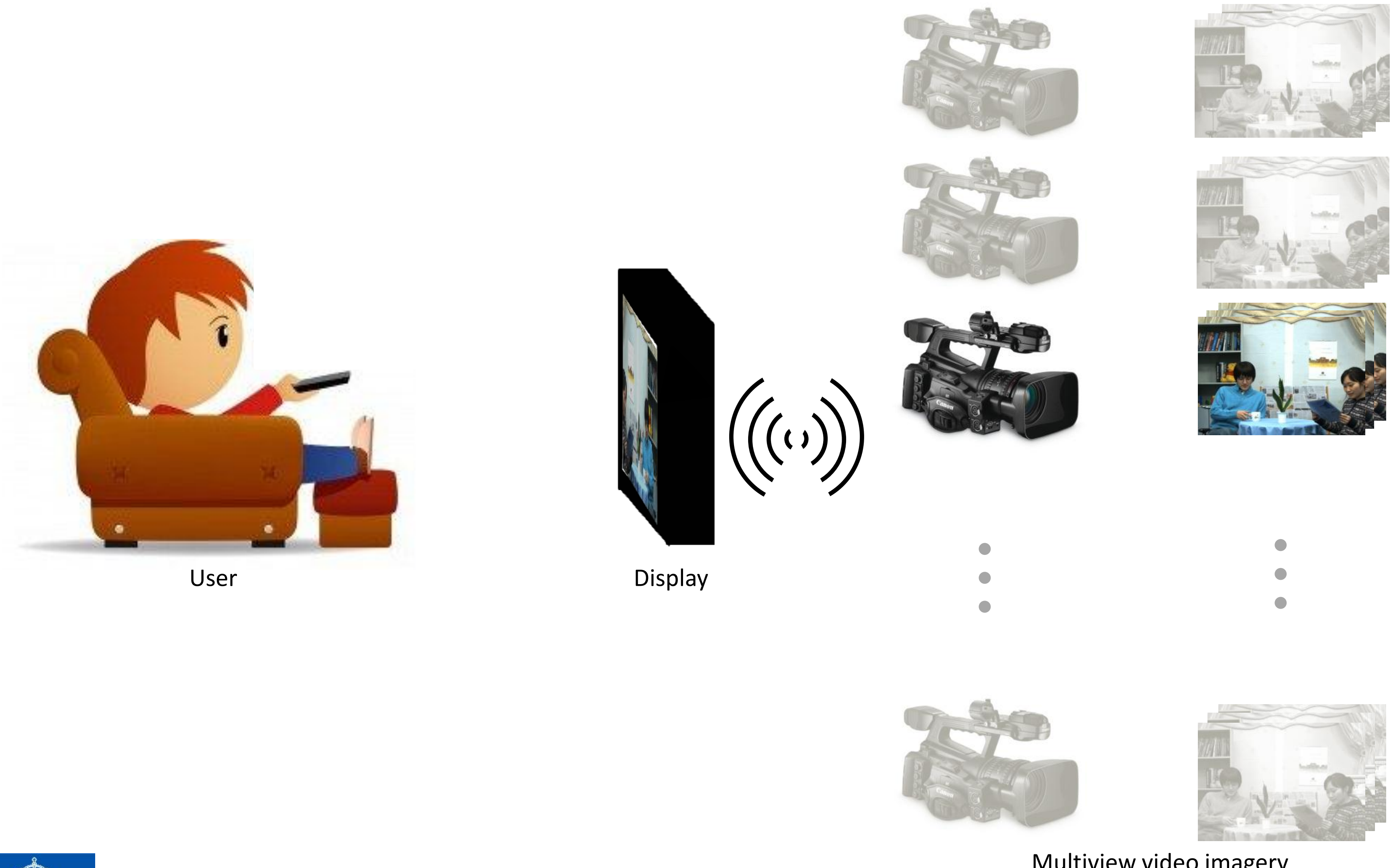
December 10, 2012

Background and motivation

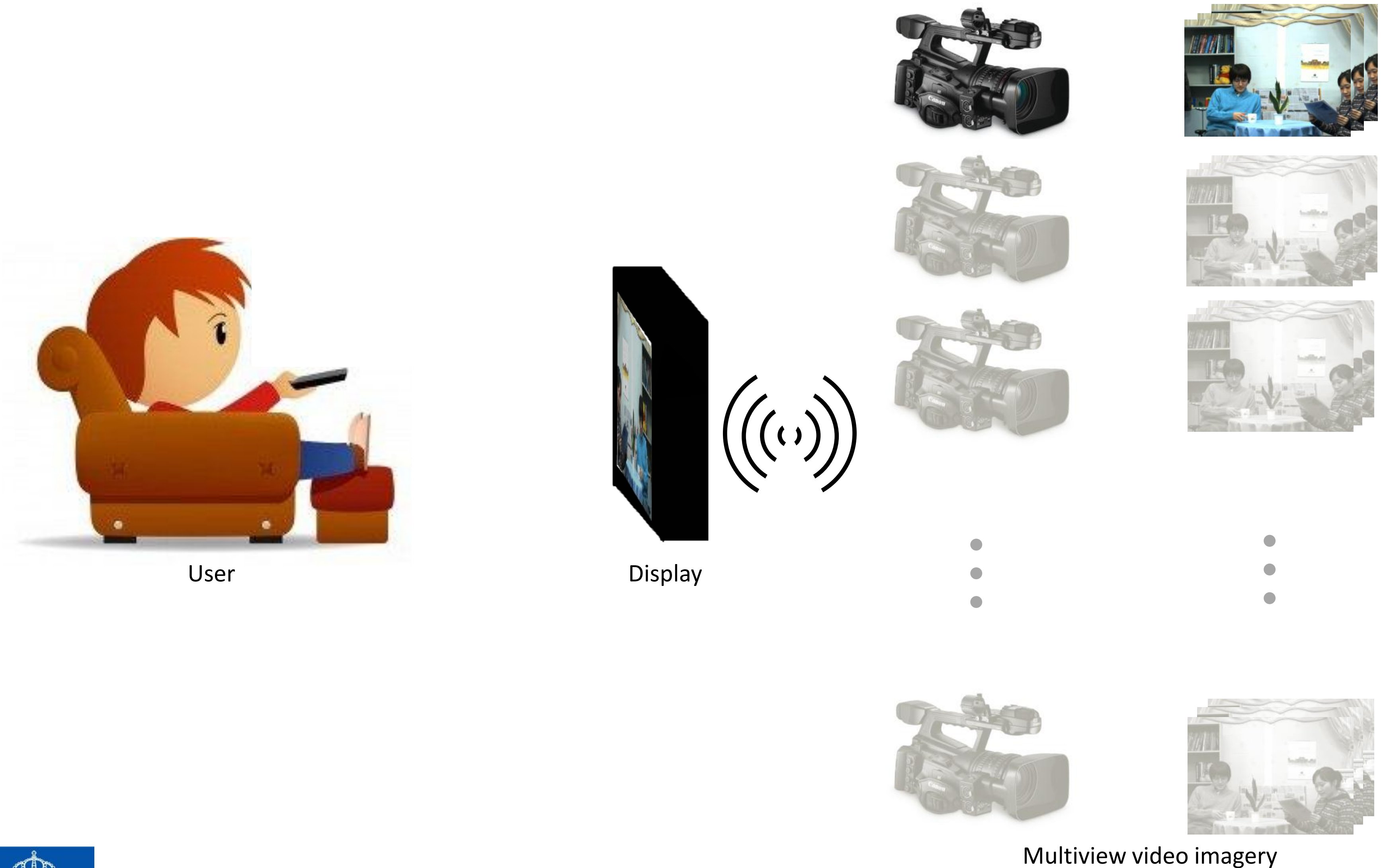
Free-viewpoint television



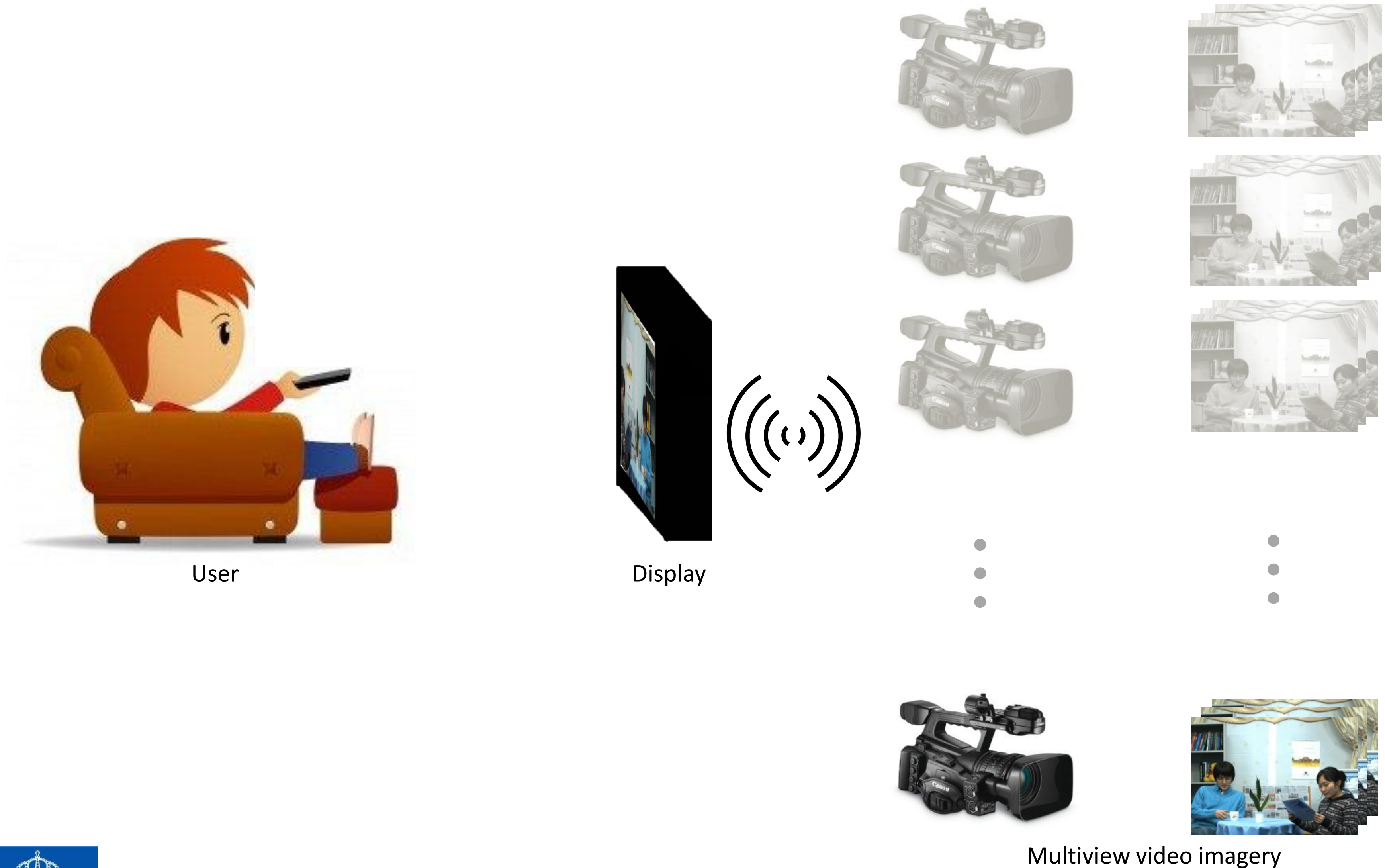
Free-viewpoint television



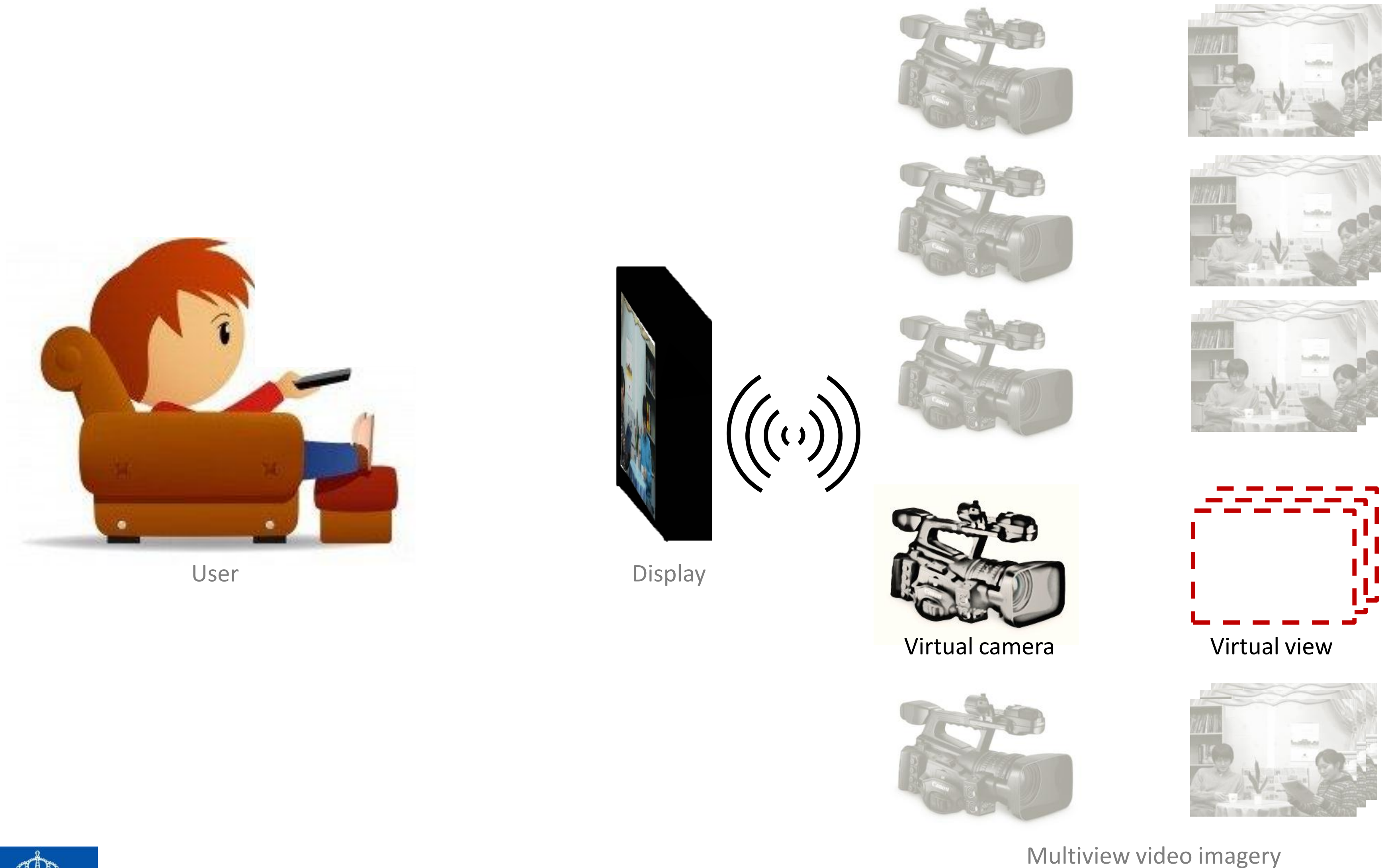
Free-viewpoint television



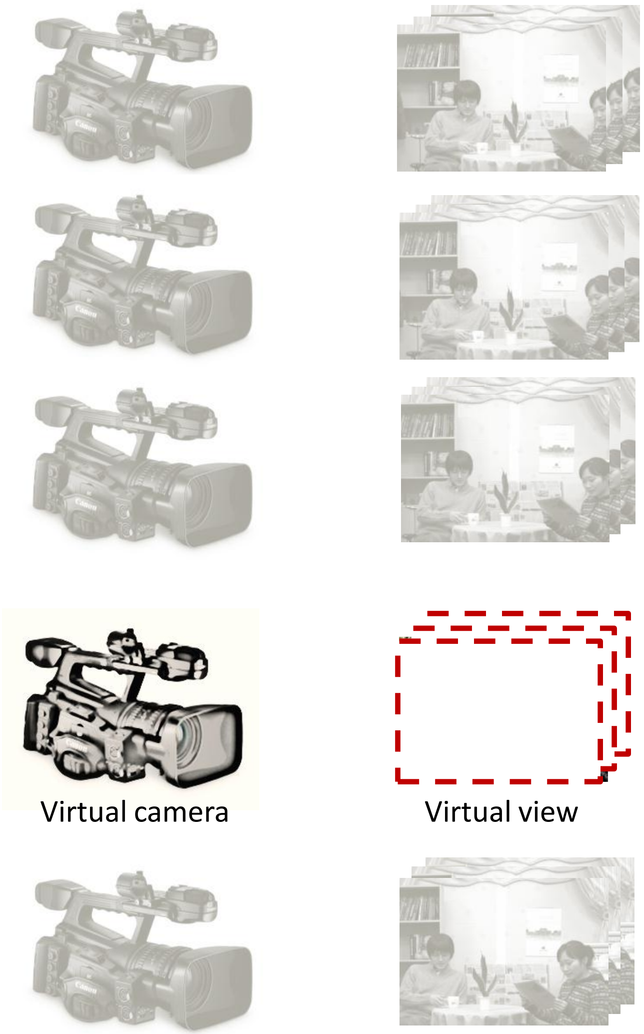
Free-viewpoint television



Free-viewpoint television

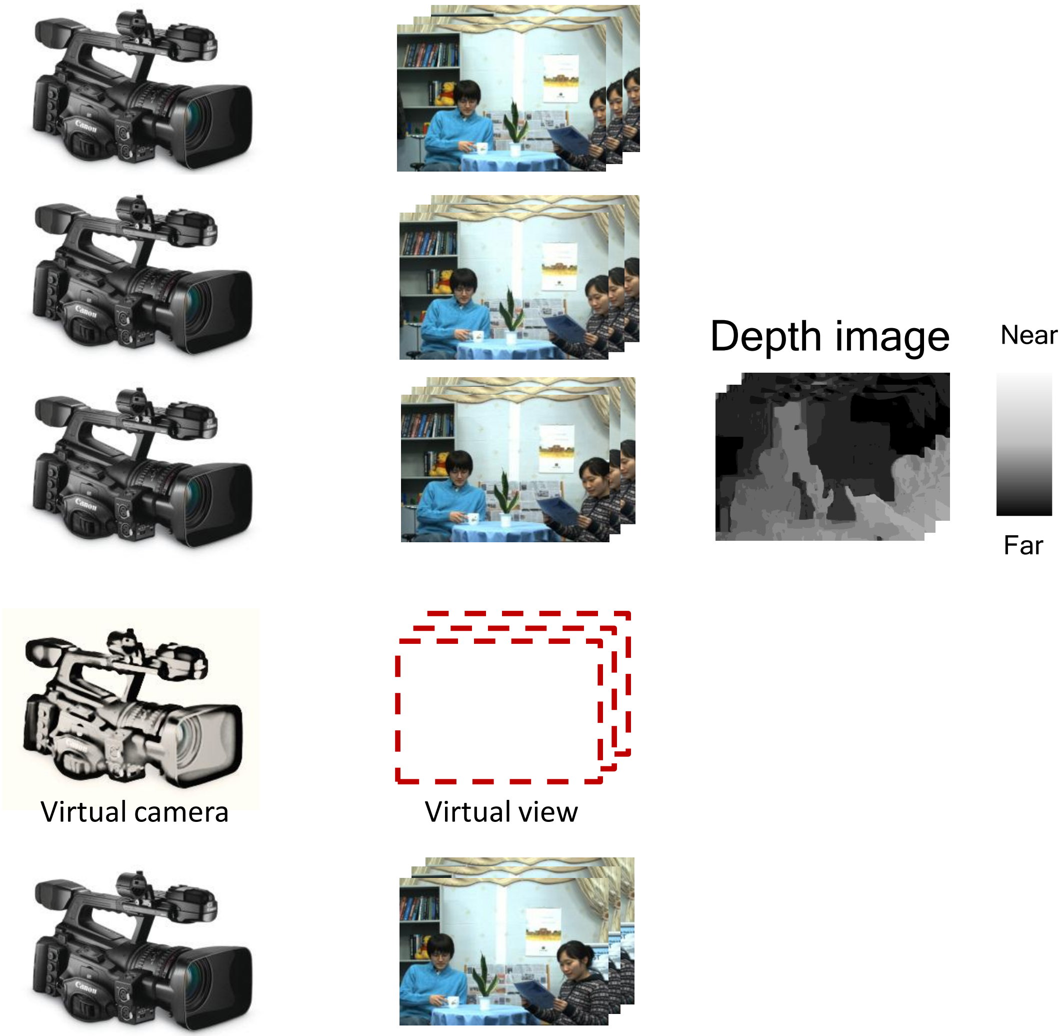


Depth image based rendering



Multiview video imagery

Depth image based rendering



Multiview video imagery

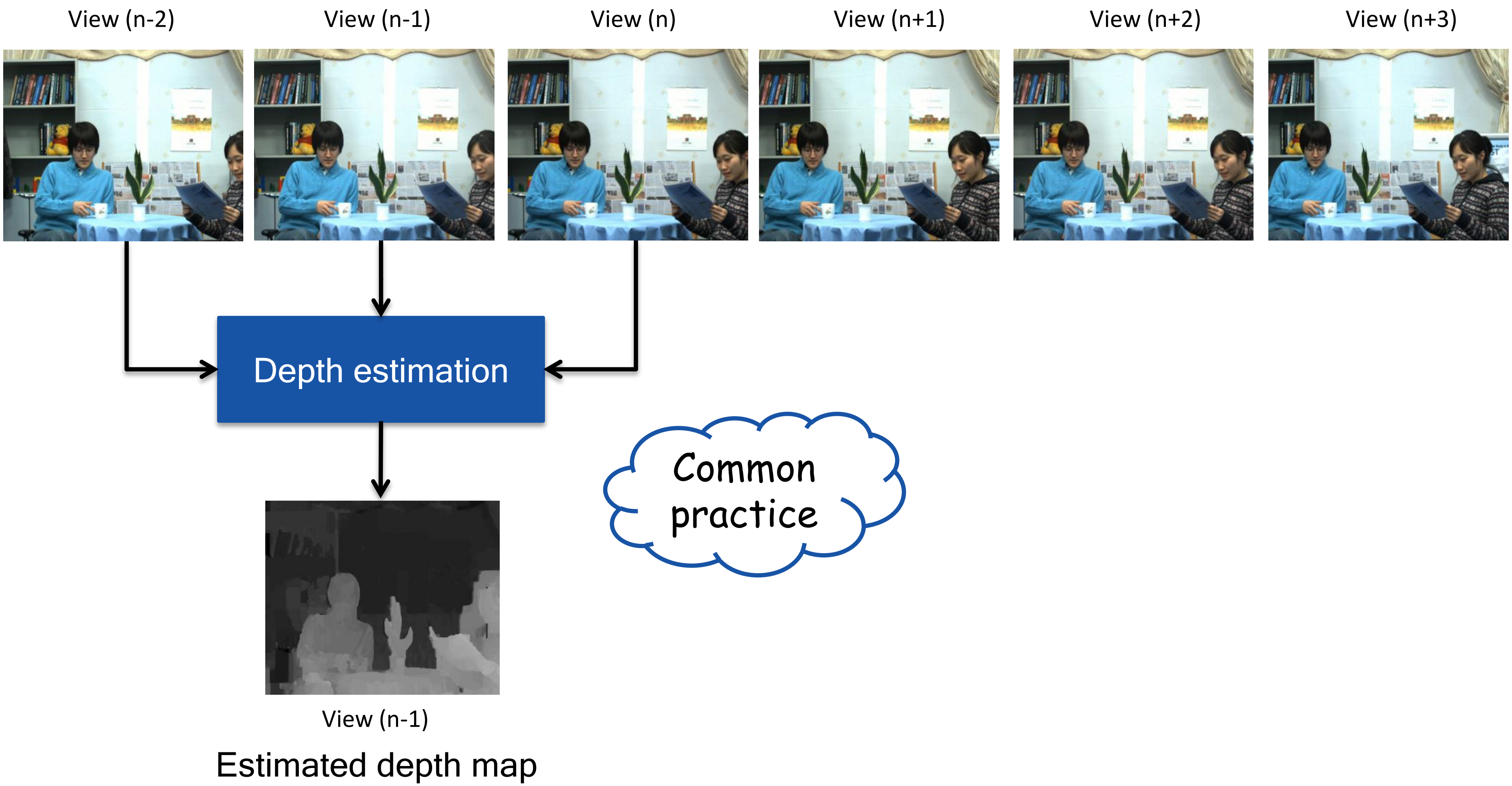
- Depth pixels represent shortest distance between object points and the camera plane
- To be estimated from multiview imagery

Depth image based rendering

- Depth pixels represent shortest distance between object points and the camera plane
- To be estimated from multiview imagery

Depth estimation

MPEG Depth Estimation Reference Software



Common practice

Depth estimation

MPEG Depth Estimation Reference Software

View (n-2) View (n-1) View (n) View (n+1) View (n+2) View (n+3)

Depth estimation

Depth estimation

Common practice

View (n-1)

Estimated depth map

View (n+2)

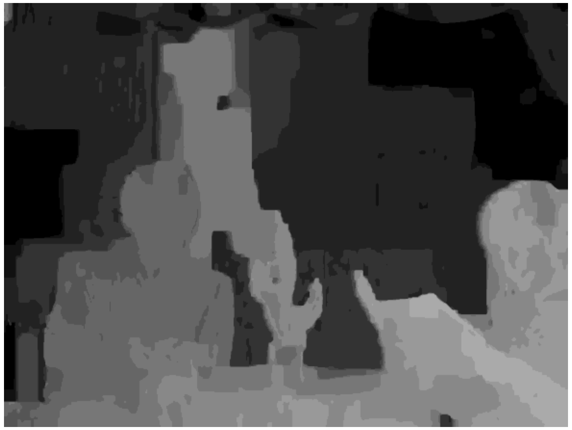
Estimated depth map

Problem: Inter-view depth inconsistency

View (n-1)

Problem: Inter-view depth inconsistency

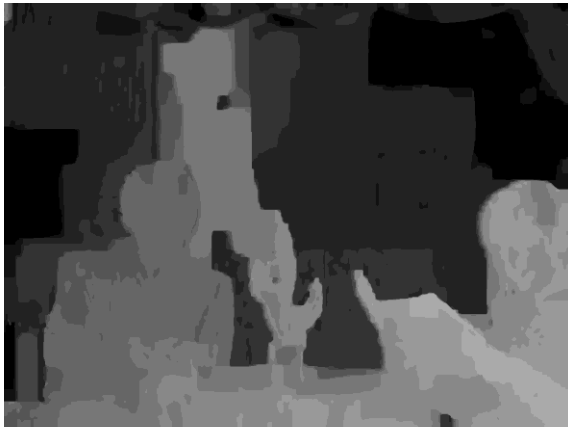
View (n-1)



View (n)

Problem: Inter-view depth inconsistency

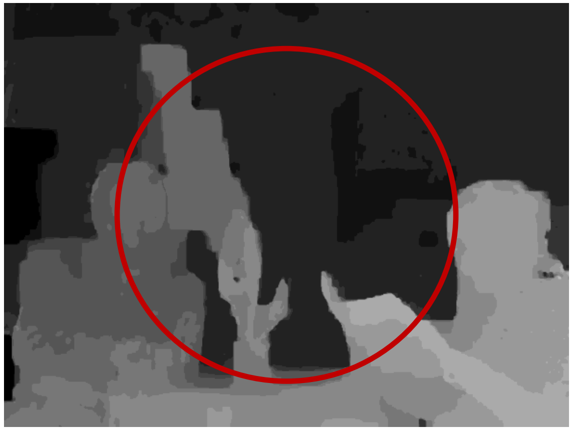
View (n-1)



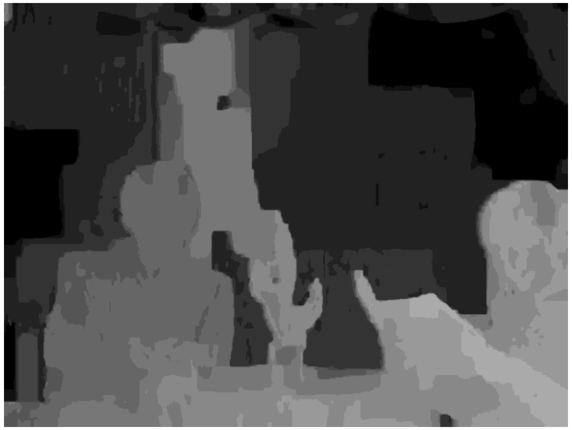
View (n)

View (n+1)

Problem: Inter-view depth inconsistency



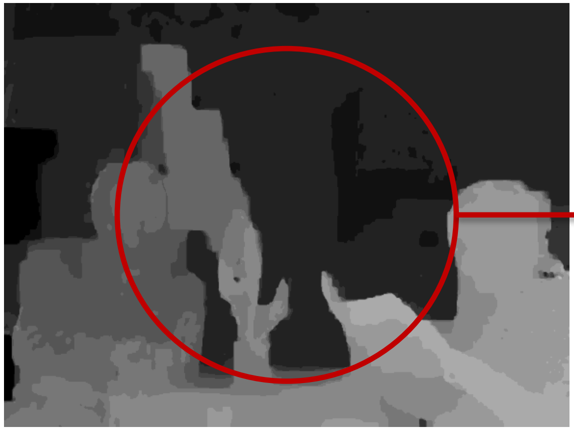
View (n-1)



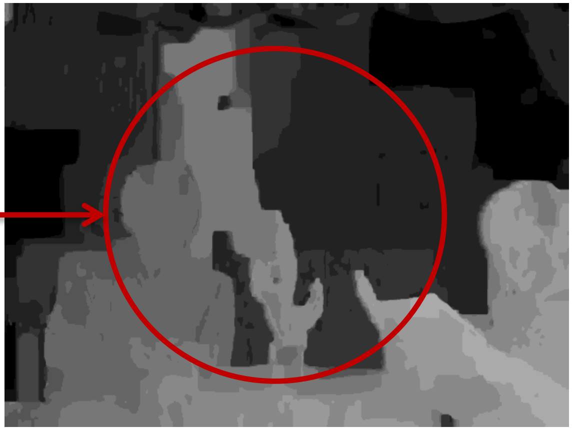
View (n)

View (n+1)

Problem: Inter-view depth inconsistency



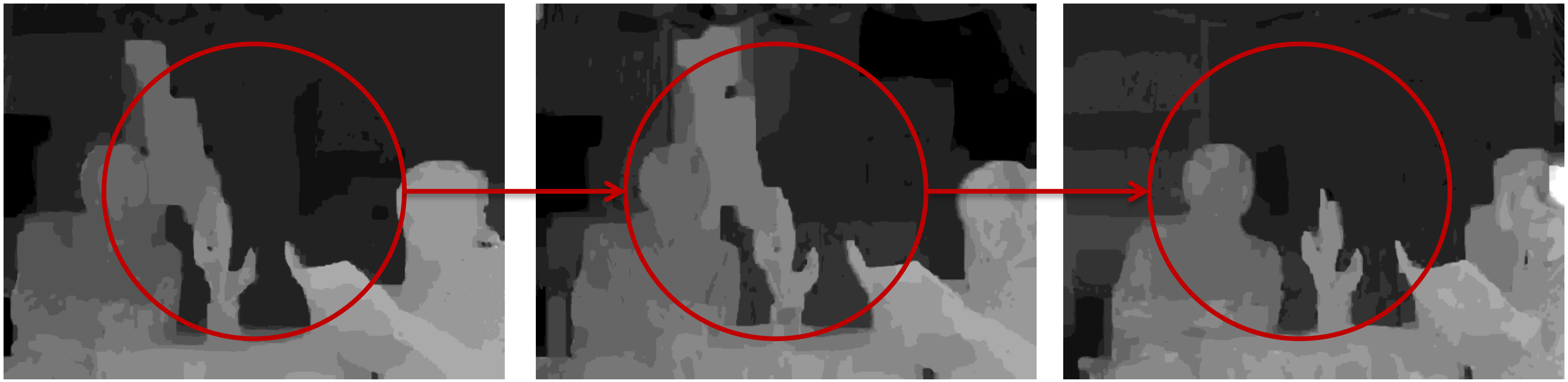
View (n-1)



View (n)

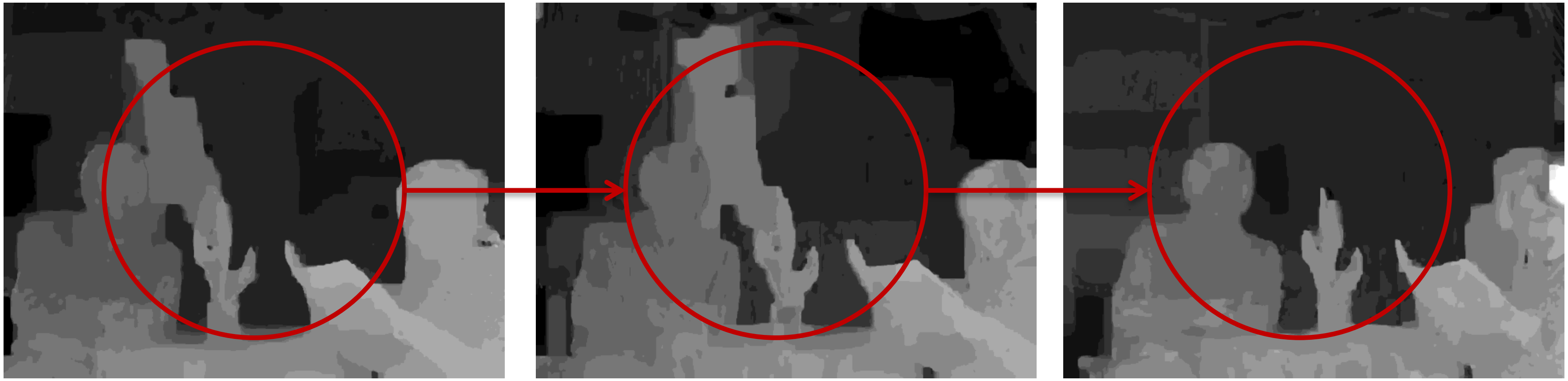
View (n+1)

Problem: Inter-view depth inconsistency



View (n-1) View (n) View (n+1)

Problem: Inter-view depth inconsistency



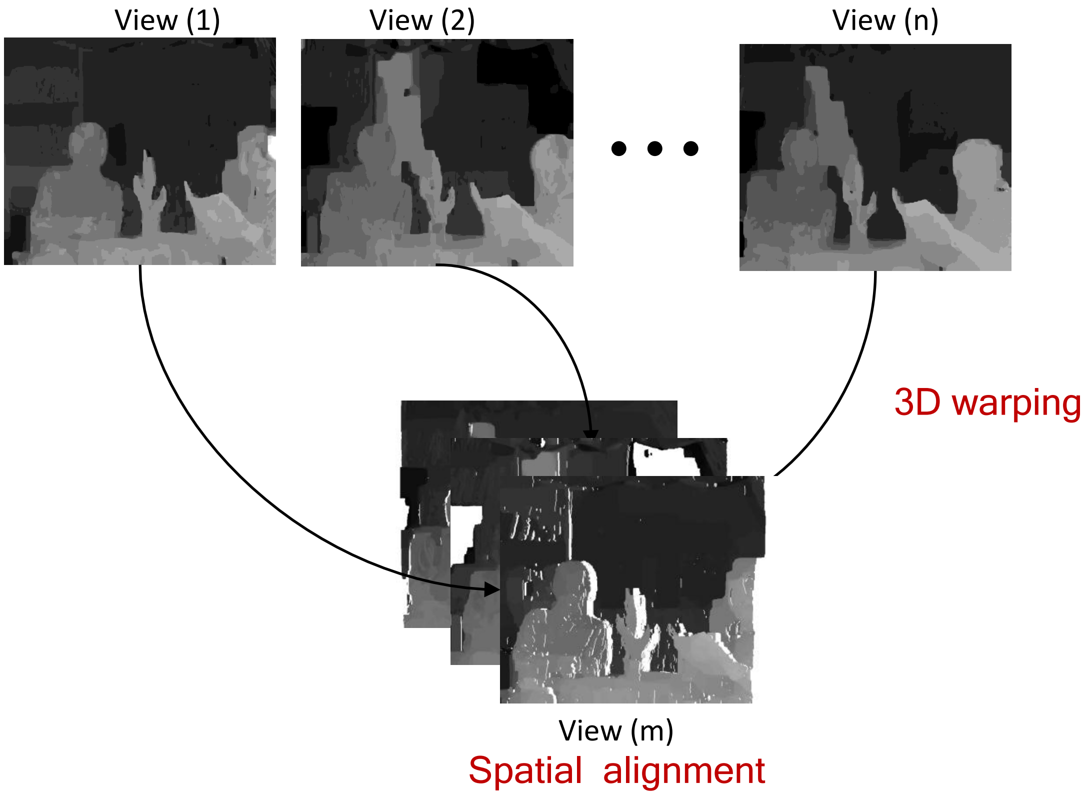
View (n-1) View (n) View (n+1)

Note: we assume a 1D-parallel camera arrangement

Prior work on depth enhancement

1. Existing methods warp depth images from multiple viewpoints to a common viewpoint for spatial alignment ([2], [3])

2. Warping errors due to the discrete values in depth maps affects enhancement algorithms negatively



[2] P. K. Rana and M. Flierl, "Depth consistency testing for improved view interpolation," IEEE Int. Workshop MMSP, 2010.

[3] E. Ekmekcioglu, V. Velisavljevic, and S. Worrall, "Content adaptive enhancement of multi-view depth maps for free viewpoint video," IEEE J. Sel. Topics Signal Process., 2011.

New depth enhancement framework

Overview of new depth enhancement framework

- Concatenation of view imagery
- Multiview color classification
- Multiview depth classification
- Depth image enhancement

Concatenation of view imagery

- The captured MVV imagery of the scene has inherent inter-view similarity
- To have a unique model for the captured natural scene,

The MVV inter-view similarity is exploited by concatenating views from multiple viewpoints

• • •

View (n-1)

View (n)

View (n+1)

View (n+2)

• • •

Concatenation of view imagery

- The captured MVV imagery of the scene has inherent inter-view similarity
- To have a unique model for the captured natural scene,

The MVV inter-view similarity is exploited by concatenating views from multiple viewpoints

View (n-1)

View (n)

View (n+1)

View (n+2)

View (n-1) View (n) View (n+1) View (n+2)

Multiview color classification

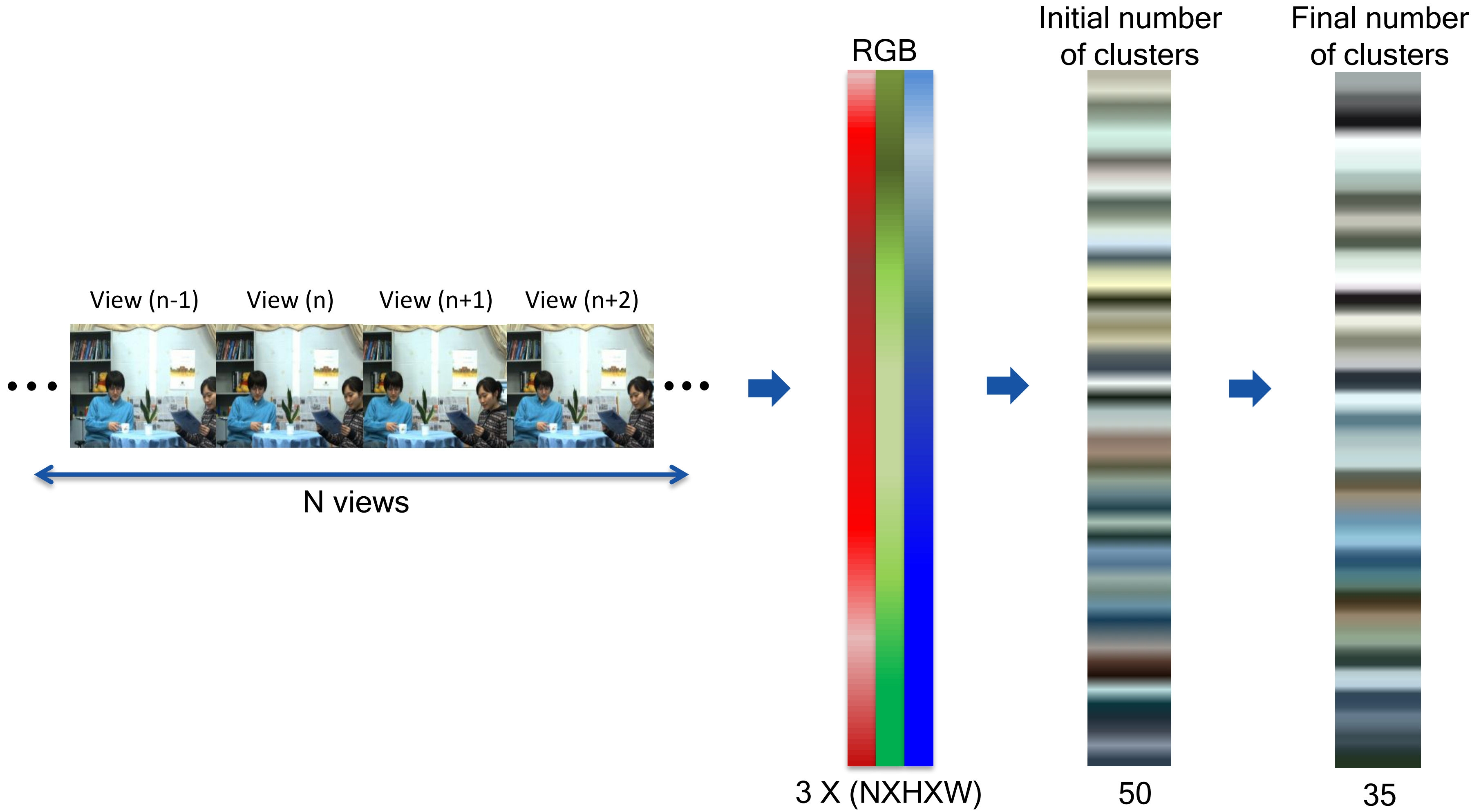
Gaussian mixture model with variational Bayes inference

- The goal of classification is to partition an image into regions each of which has a reasonably homogeneous visual appearance
- Usually, classification algorithm, such as expectation-maximization for Gaussian mixtures, suffers from two main drawbacks:
  - model over-fitting and
  - the number of clusters has to be known, (similar to the K-means algorithm)
- The Gaussian mixture model is used with variational Bayes inference [4] because
  - no model over-fitting and
  - the number of clusters is treated as a random variable

[4] C. M. Bishop, Pattern Recognition and Machine Learning, 1st ed. New York: Springer, 2006.

Multiview color classification

Gaussian mixture model with variational Bayes inference



[4] C. M. Bishop, Pattern Recognition and Machine Learning, 1st ed. New York: Springer, 2006.

Multiview color classification

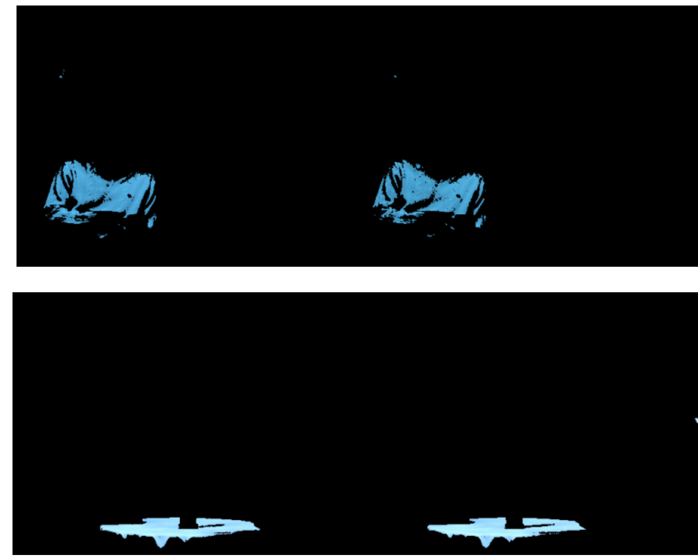
Example: Newspaper

Color classification input

Color clusters

Multiview color classification

Example: Newspaper



Color classification input

Color clusters

Multiview color classification

Example: Newspaper

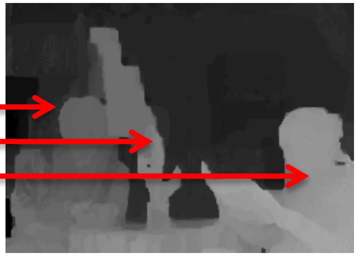
Color classification input

Color clusters

Multiview depth classification

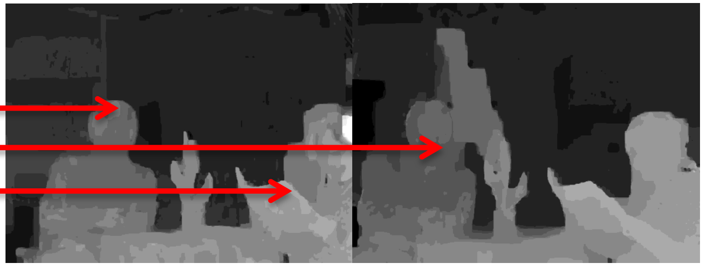
Exploiting the per-pixel association between color and depth

View image



Depth image

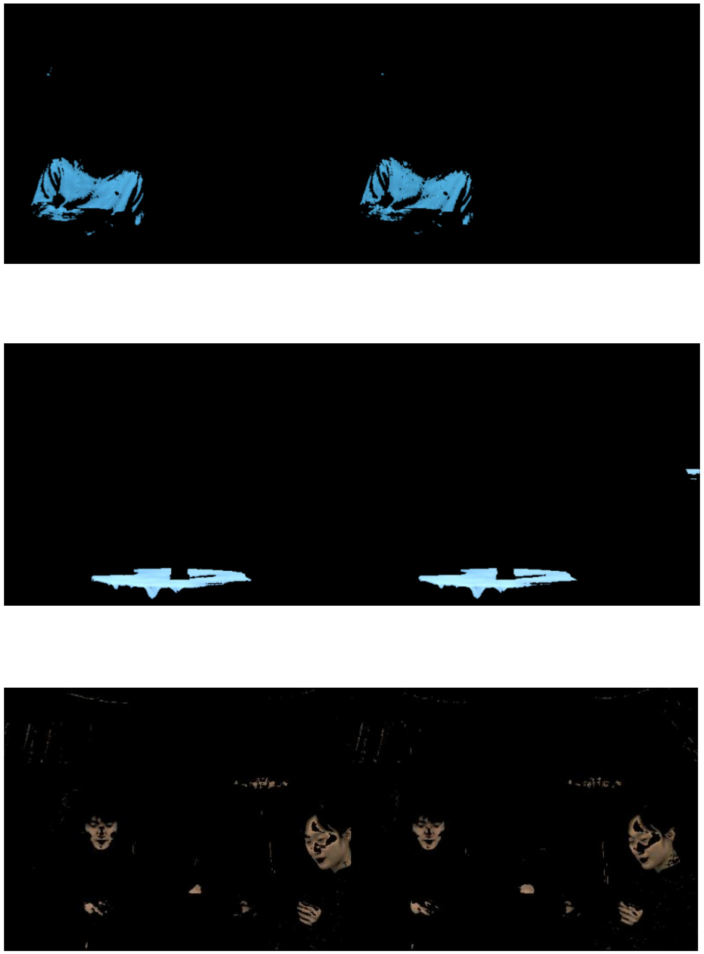
Concatenated view imagery



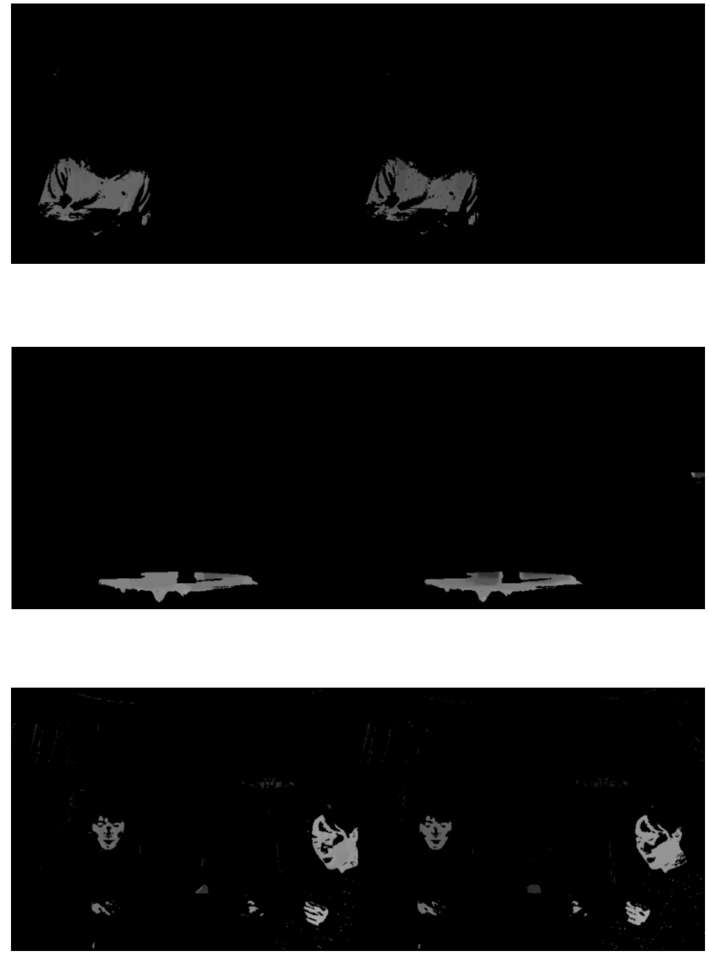
Concatenated depth imagery

Multiview depth classification

Example: Newspaper



Color clusters



Depth clusters

Multiview depth enhancement

Difference between color and depth clusters

Members have similar colors pixels

Members may have different depth values

Multiview depth enhancement

Difference between color and depth clusters

Members have similar colors pixels

Members may have different depth values

- Why?
  - Due to foreground and background depth difference
  - Due to inter-view inconsistency

Multiview depth enhancement

Difference between color and depth clusters

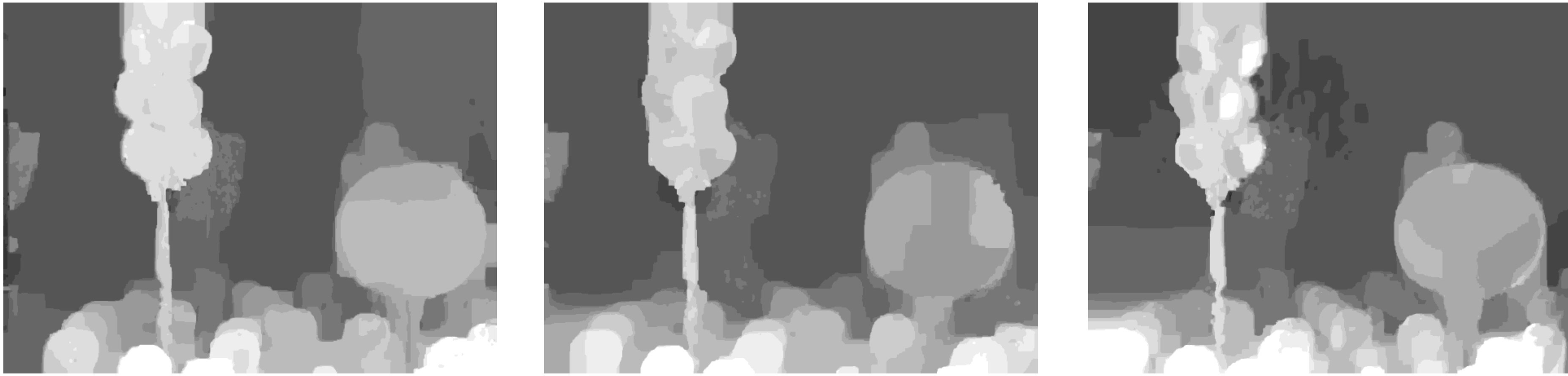
Members have similar colors pixels

Members may have different depth values

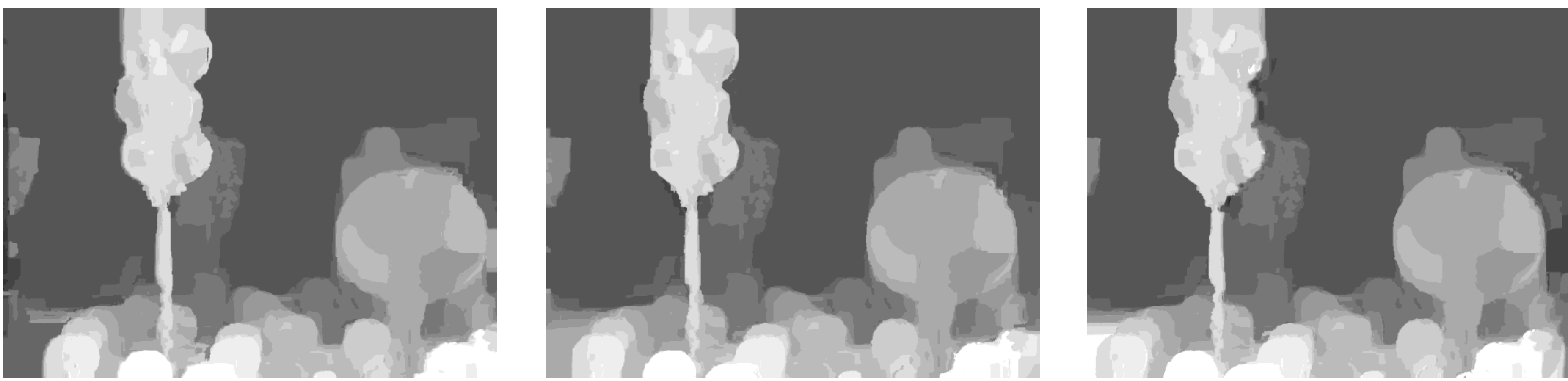
- Why?
  - Due to foreground and background depth difference
  - Due to inter-view inconsistency
- Our approach: K-means sub-clustering
  - Computationally fast
  - Assigns the mean to depth pixels irrespective of the originating viewpoints
  - Usually, Bayesian approaches imply higher computational complexity

Multiview depth enhancement

Example: Balloons



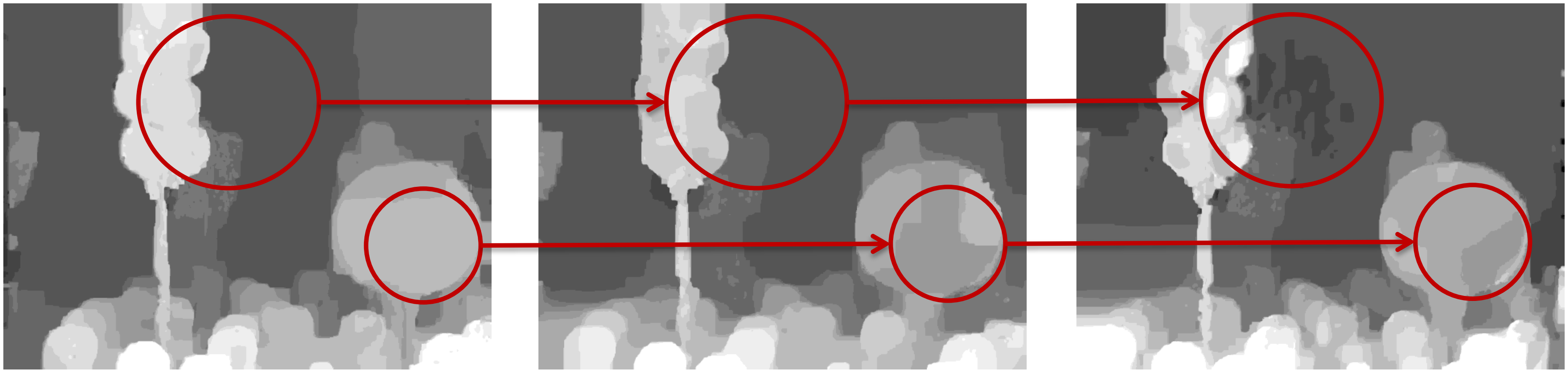
MPEG depth maps



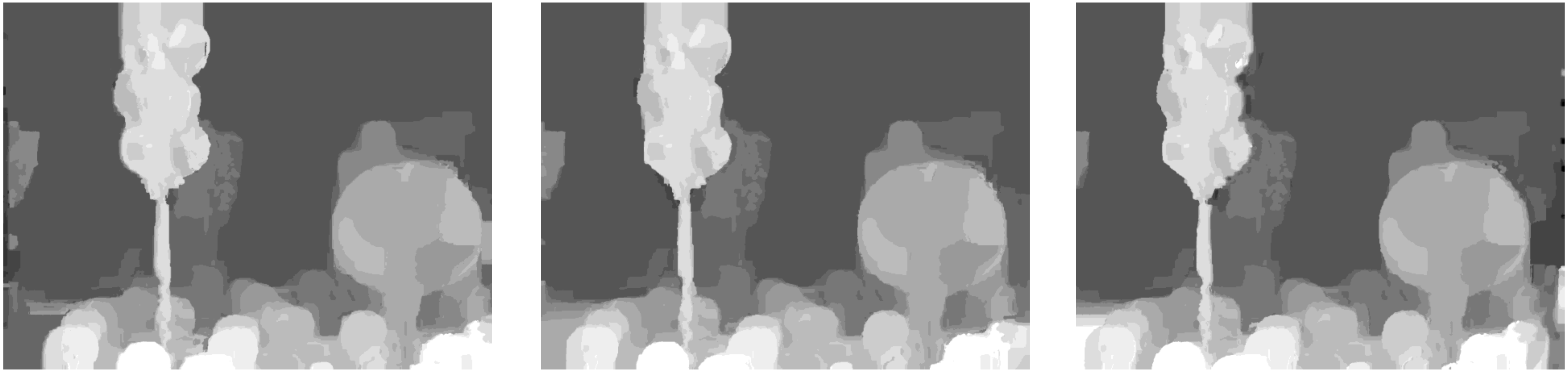
Enhanced depth maps

Multiview depth enhancement

Example: Balloons



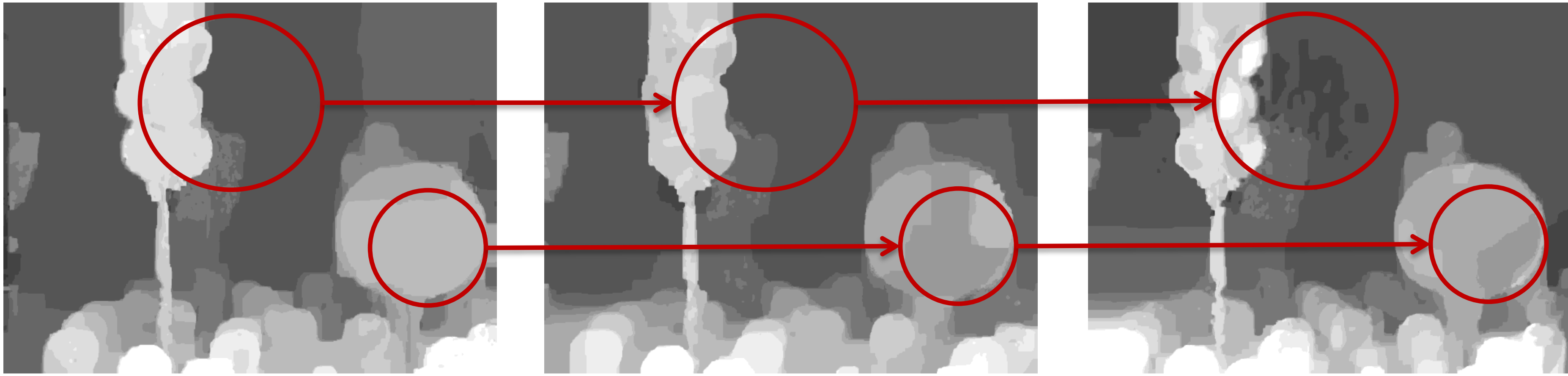
MPEG depth maps



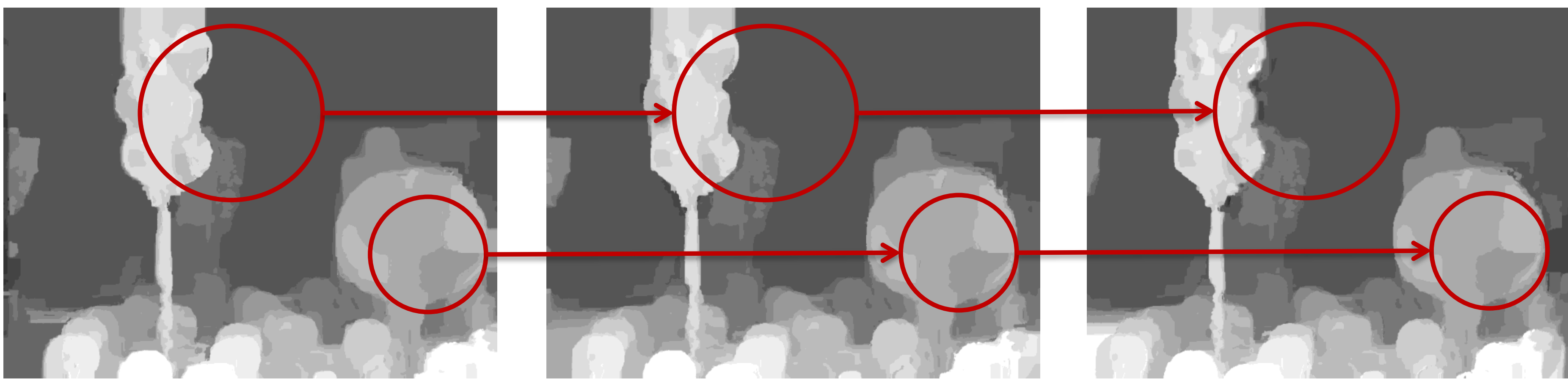
Enhanced depth maps

Multiview depth enhancement

Example: Balloons



MPEG depth maps



Enhanced depth maps

Experimental setup

MPEG 3DTV multiview data set

Newspaper
(1024 X 768)

Lovebird1
(1024 X 768)

Kendo
(1024 X 768)

Balloons
(1024 X 768)

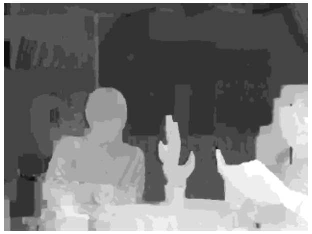
Poznan street
(1920 X 1088)

Depth image based rendering

MPEG View Synthesis Reference Software (VSRS) 3.5

Enhanced depth map

Left



Enhanced depth map

Right

Depth image based rendering

MPEG View Synthesis Reference Software (VSRS) 3.5

Left

Enhanced depth map

Reference view

3D warping

Right

Enhanced depth map

Reference view

3D warping

Depth image based rendering

MPEG View Synthesis Reference Software (VSRS) 3.5

Left

Enhanced depth map

Reference view

3D warping

Warped view

Right

Enhanced depth map

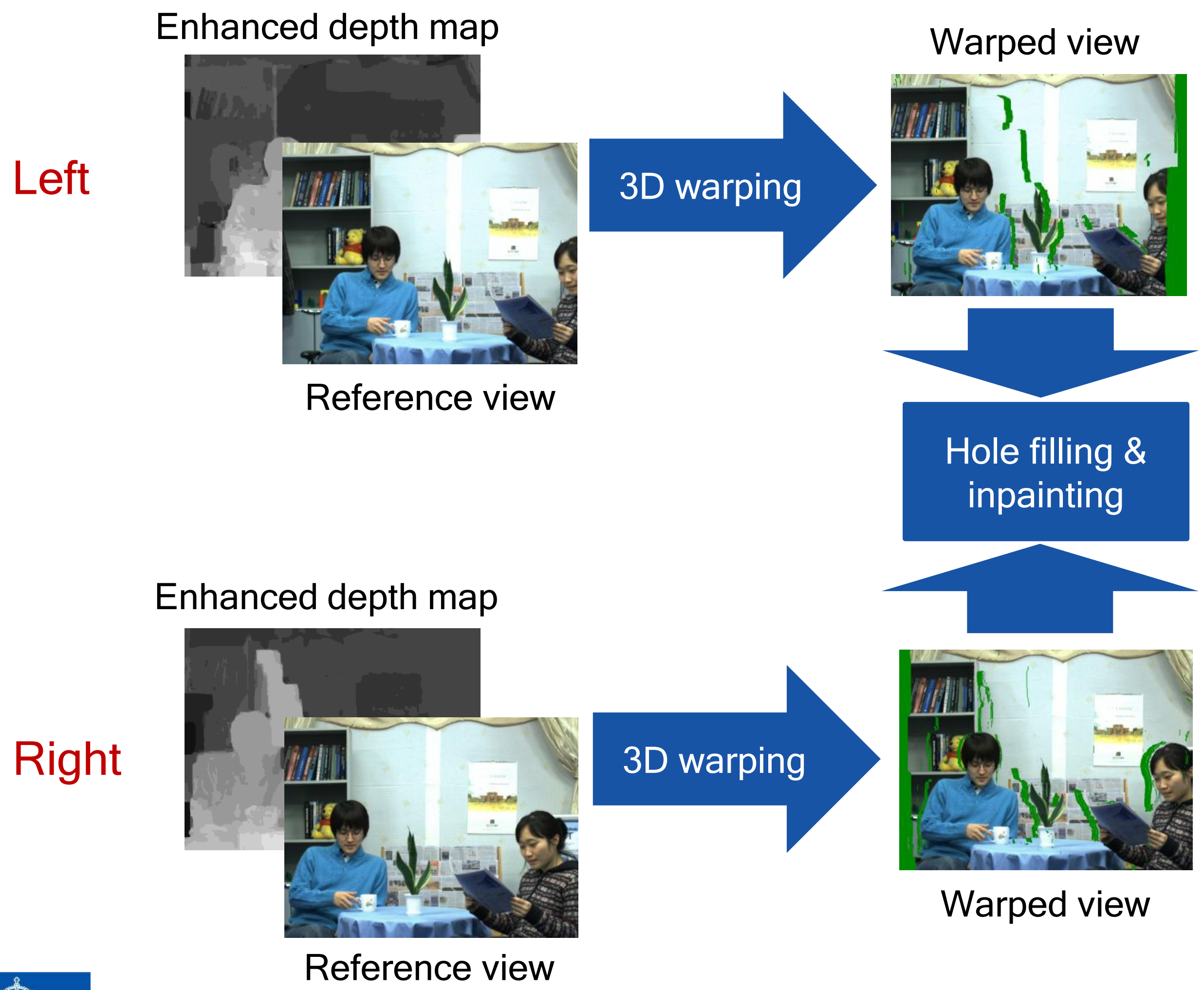
Reference view

3D warping

Warped view

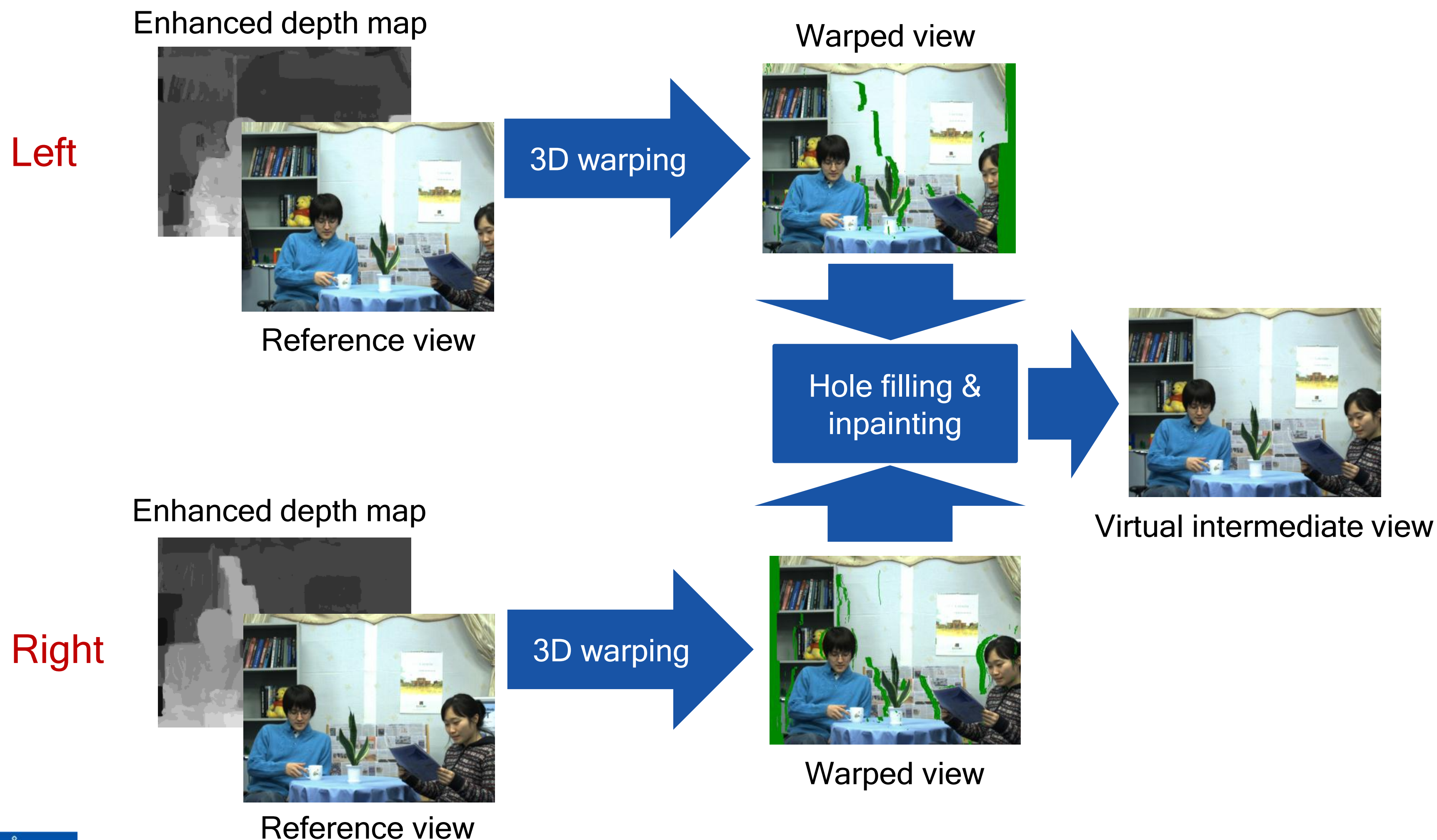
Depth image based rendering

MPEG View Synthesis Reference Software (VSRS) 3.5



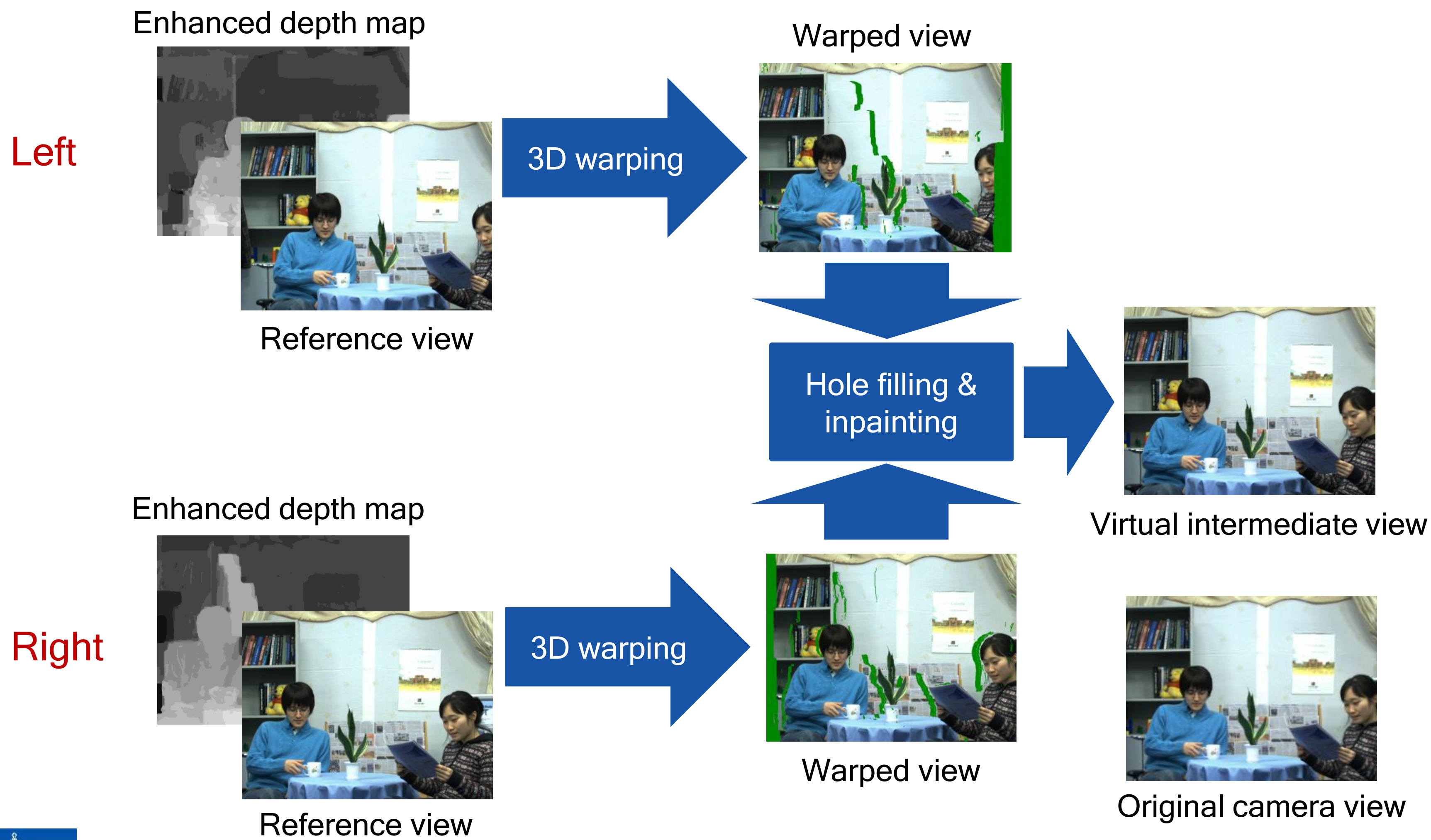
Depth image based rendering

MPEG View Synthesis Reference Software (VSRS) 3.5



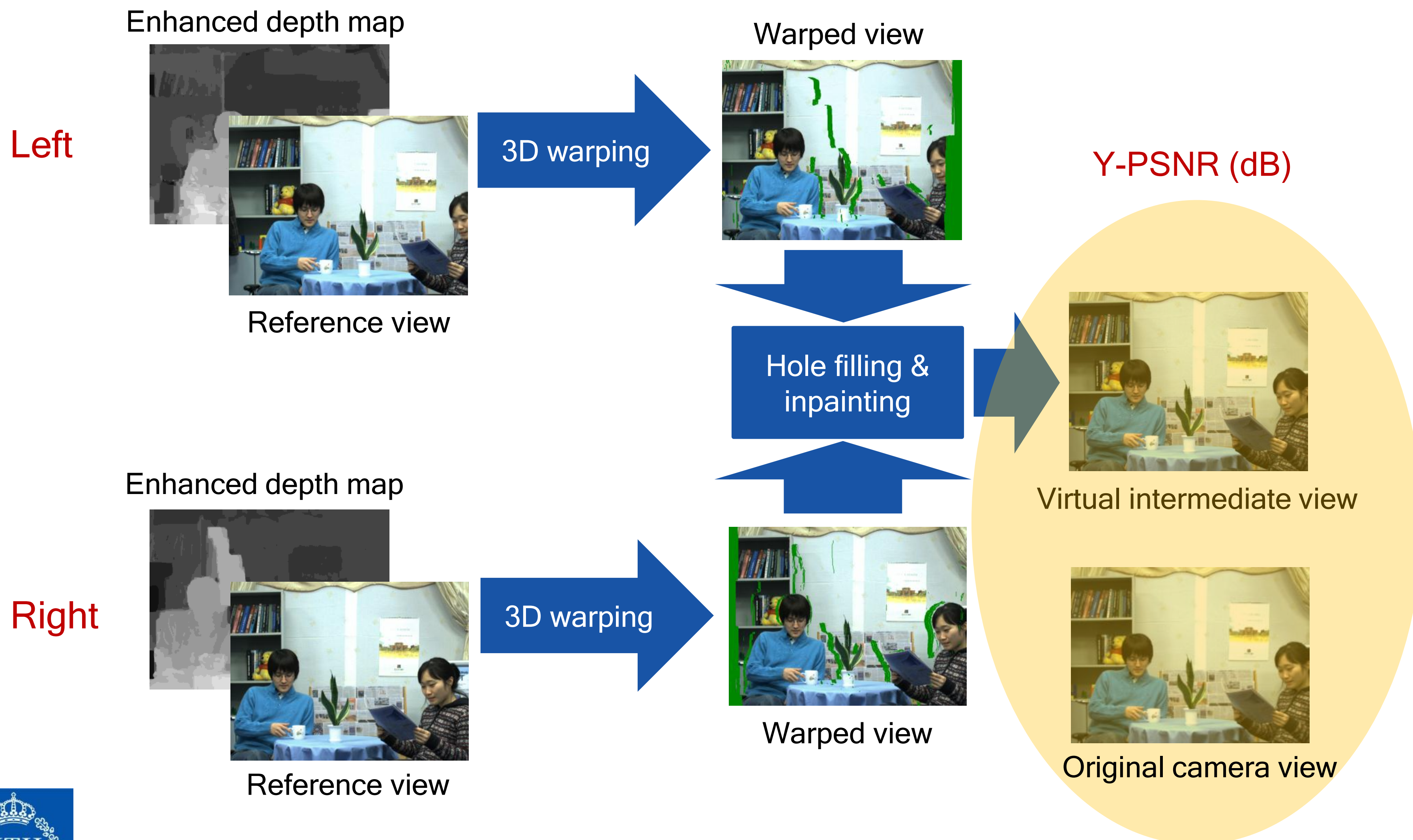
Depth image based rendering

MPEG View Synthesis Reference Software (VSRS) 3.5



Depth image based rendering

MPEG View Synthesis Reference Software (VSRS) 3.5



Objective results

Test sequence	Sequence resolution	Input views	Virtual view	MPEG VSRS views Y-PSNR 3.5 [dB]	
				MPEG depth maps	Enhanced depth maps
Newspaper	1024 X 768	4,6	5	31.98	32.10
Kendo	1024 X 768	3,5	5	36.54	36.72
Poznan Street	1920 X 1088	3,5	4	35.56	35.58
Lovebird1	1024 X 768	6,8	7	28.50	28.68
Balloons	1024 X 768	3,5	4	35.68	35.93

- Color classification
  - Initial number of color clusters: 50
- K-means sub-clustering
  - Number of cluster : 12

Objective results

Test sequence	Sequence resolution	Input views	Virtual view	MPEG VSRS views Y-PSNR 3.5 [dB]	
				MPEG depth maps	Enhanced depth maps
Newspaper	1024 X 768	4,6	5	31.98	32.10
Kendo	1024 X 768	3,5	5	36.54	36.72
Poznan Street	1920 X 1088	3,5	4	35.56	35.58
Lovebird1	1024 X 768	6,8	7	28.50	28.68
Balloons	1024 X 768	3,5	4	35.68	35.93

- Color classification
  - Initial number of color clusters: 50
- K-means sub-clustering
  - Number of cluster : 12

Objective results

Test sequence	Sequence resolution	Input views	Virtual view	MPEG VSRS views Y-PSNR 3.5 [dB]	
				MPEG depth maps	Enhanced depth maps
Newspaper	1024 X 768	4,6	5	31.98	32.10
Kendo	1024 X 768	3,5	5	36.54	36.72
Poznan Street	1920 X 1088	3,5	4	35.56	35.58
Lovebird1	1024 X 768	6,8	7	28.50	28.68
Balloons	1024 X 768	3,5	4	35.68	35.93

- Color classification
  - Initial number of color clusters: 50
- K-means sub-clustering
  - Number of cluster : 12

Objective results

Test sequence	Sequence resolution	Input views	Virtual view	MPEG VSRS 3.5 [dB]	
				MPEG depth maps	Enhanced depth maps
Newspaper	1024 X 768	4,6	5	31.98	32.10
Kendo	1024 X 768	3,5	5	36.54	36.72
Poznan Street	1920 X 1088	3,5	4	35.56	35.58
Lovebird1	1024 X 768	6,8	7	28.50	28.68
Balloons	1024 X 768	3,5	4	35.68	35.93

- Color classification
  - Initial number of color clusters: 50
- K-means sub-clustering
  - Number of cluster : 12

Subjective comparison

Sequence: Newspaper

With MPEG depth

With enhanced depth

Subjective comparison

Sequence: Newspaper

With MPEG depth

With enhanced depth

Subjective comparison

Sequence: Newspaper

Original

With MPEG depth

With enhanced depth

Subjective comparison

Sequence: Kendo

With MPEG depth

With enhanced depth

Subjective comparison

Sequence: Kendo

With MPEG depth

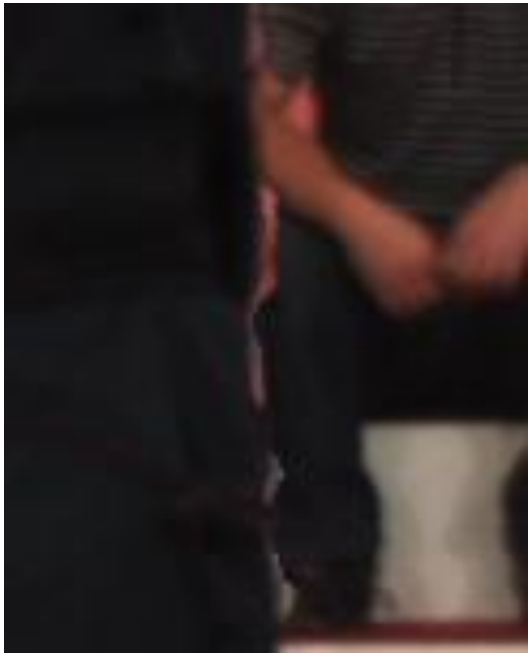
With enhanced depth

Subjective comparison

Sequence: Kendo

Original

With MPEG depth



With enhanced depth

Subjective comparison

Sequence: Lovebird 1

With MPEG depth

With enhanced depth

Subjective comparison

Sequence: Lovebird 1

With MPEG depth

With enhanced depth

Subjective comparison

Sequence: Lovebird 1

Original

With MPEG depth

With enhanced depth

Subjective comparison

Sequence: Lovebird 1

With MPEG depth

With enhanced depth

Subjective comparison

Sequence: Lovebird 1

Original

With MPEG depth

With enhanced depth

Subjective comparison

Sequence: Balloons

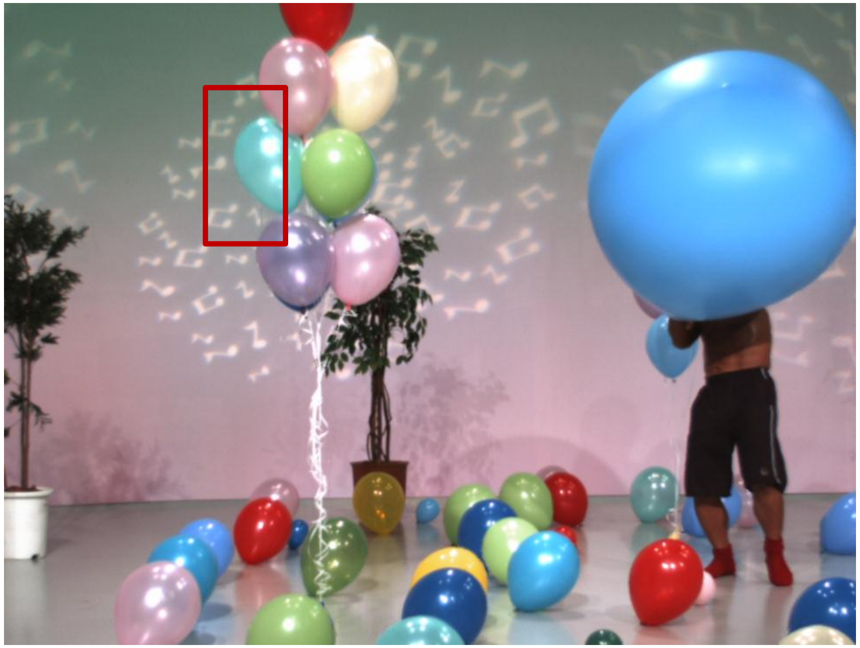
With MPEG depth

With enhanced depth

Subjective comparison

Sequence: Balloons

With MPEG depth



With enhanced depth

Subjective comparison

Sequence: Balloons

Original

With MPEG depth

With enhanced depth

Subjective comparison

Sequence: Poznan Street

With MPEG depth

With enhanced depth

Subjective comparison

Sequence: Poznan Street

With MPEG depth

With enhanced depth

Subjective comparison

Sequence: Poznan Street

Original

With MPEG depth

With enhanced depth

Conclusions

- We improved the inter-view depth consistency and hence, enhanced the visual experience of free-viewpoint television
- For that, we exploited the per-pixel association between depth and color by classification
- Color classification is accomplished by variational Bayesian inference
- Then, color classes are used for depth classification
- Effectiveness of our approach is demonstrated by both objective and subjective results

Future directions

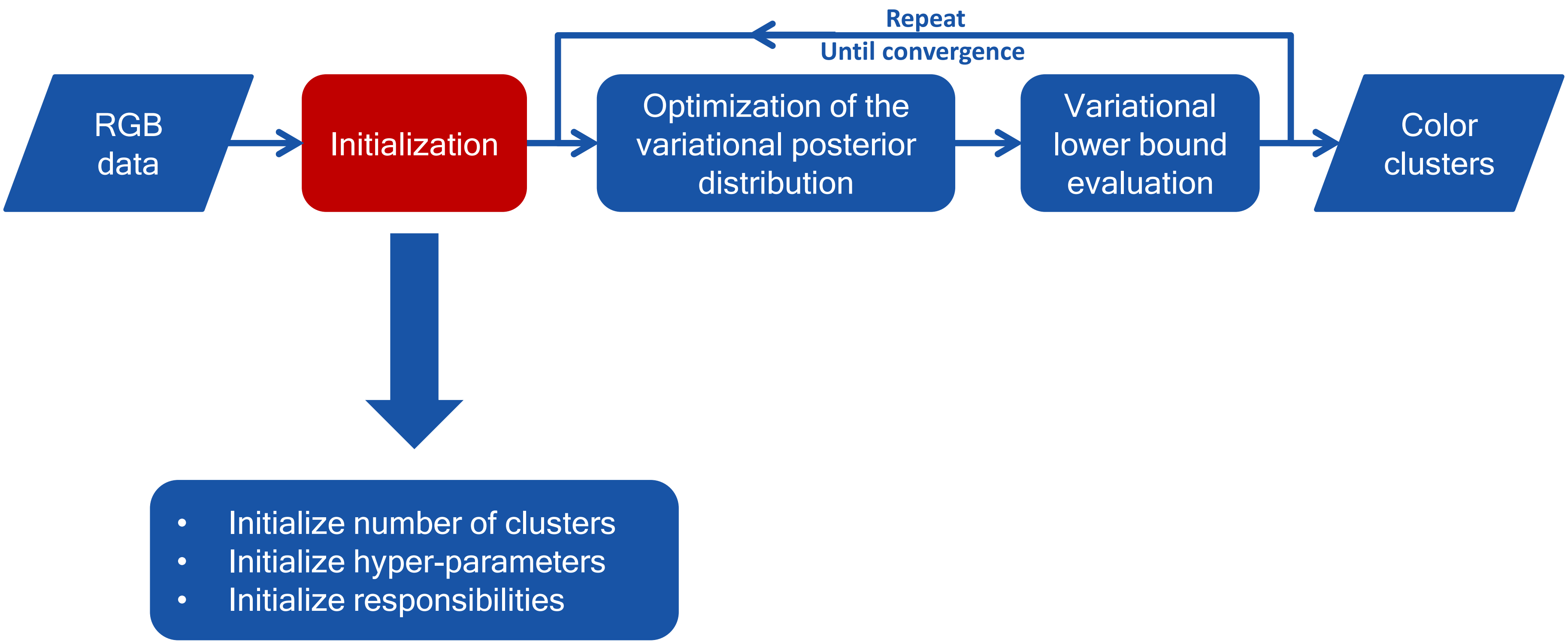
- Improve temporal depth consistency
- Improve color classification by using other mixture models
- Improve computational efficiency of color classification

Thank you

Gaussian mixture model with variational Bayes inference

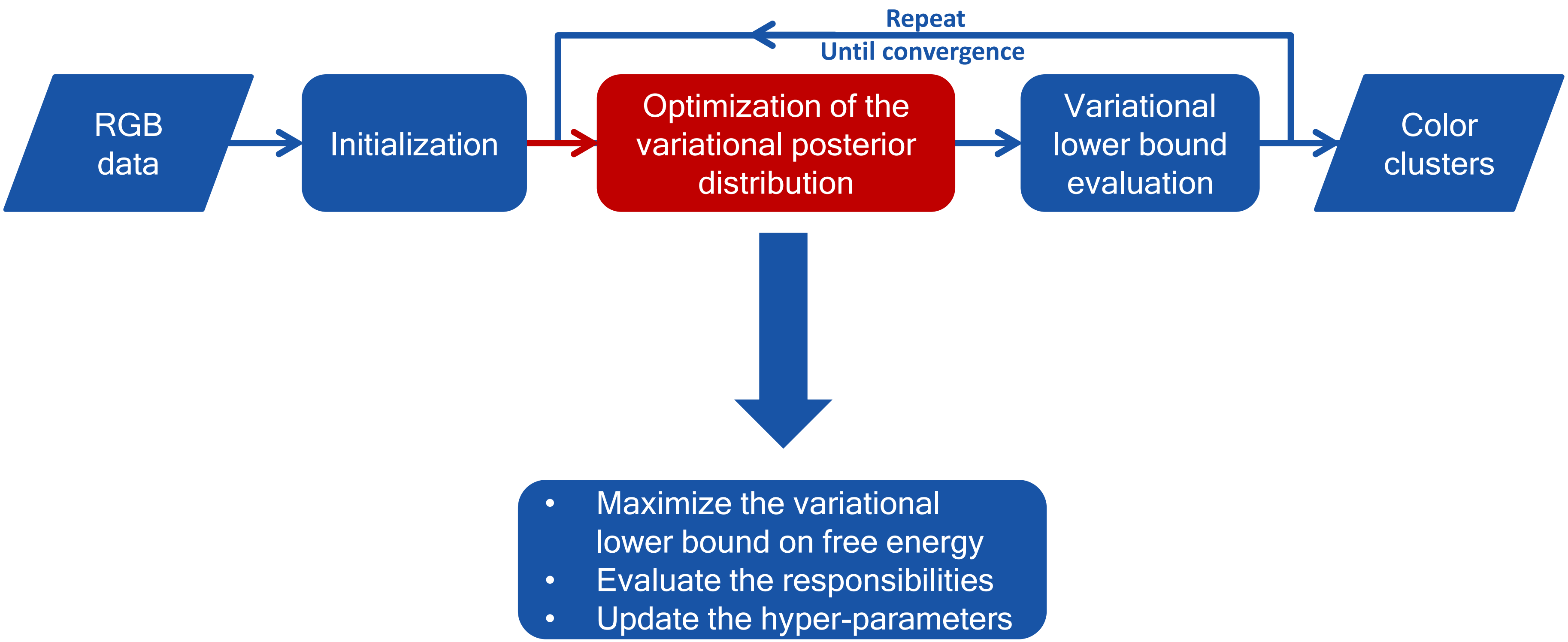
[4] C. M. Bishop, Pattern Recognition and Machine Learning, 1st ed. New York: Springer, 2006.

Gaussian mixture model with variational Bayes inference



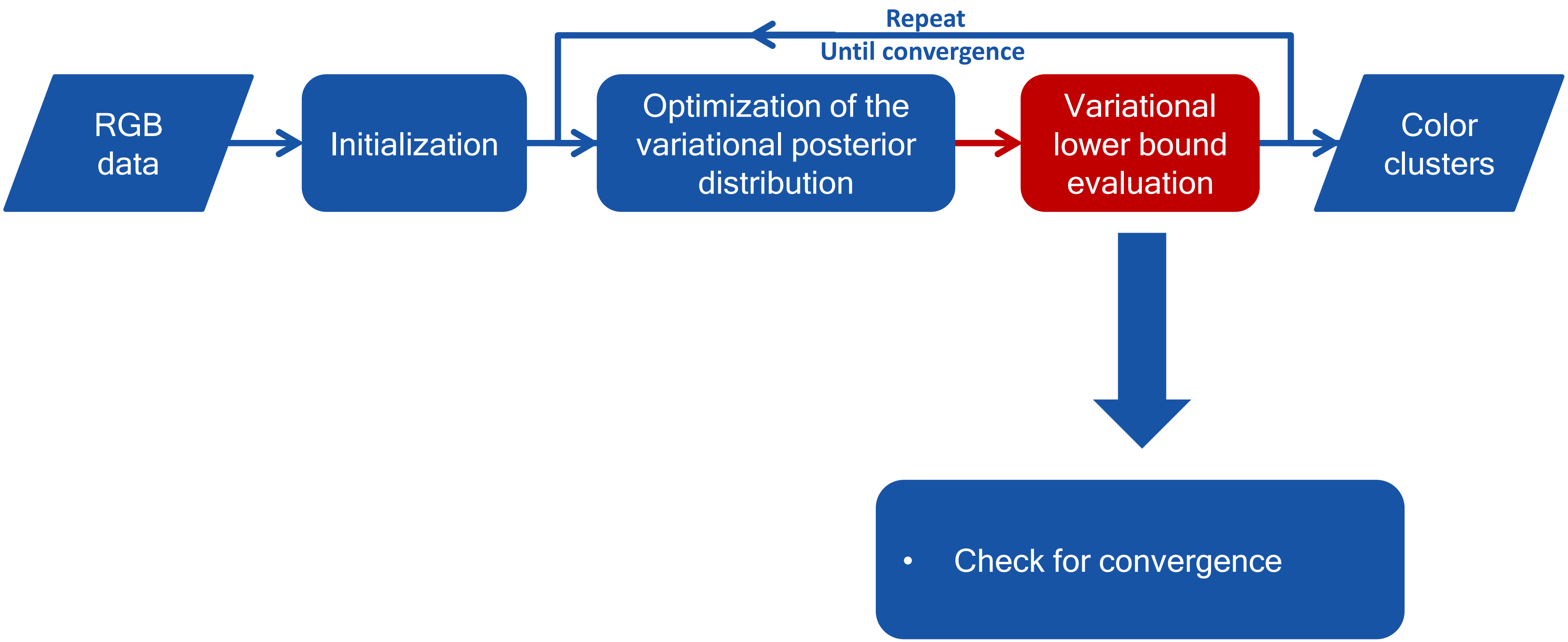
[4] C. M. Bishop, Pattern Recognition and Machine Learning, 1st ed. New York: Springer, 2006.

Gaussian mixture model with variational Bayes inference



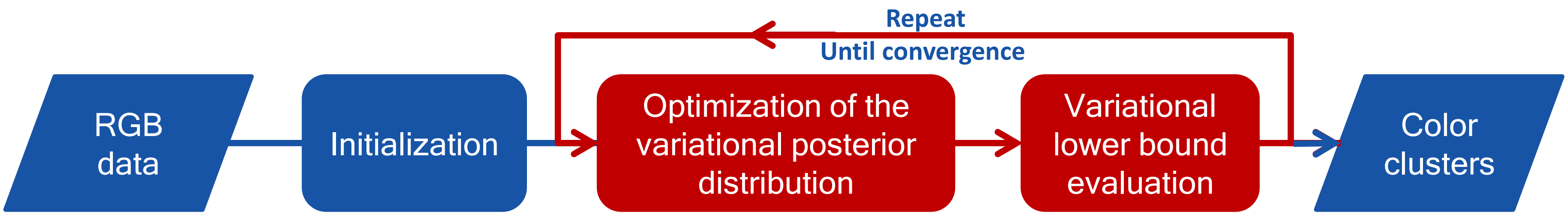
[4] C. M. Bishop, Pattern Recognition and Machine Learning, 1st ed. New York: Springer, 2006.

Gaussian mixture model with variational Bayes inference



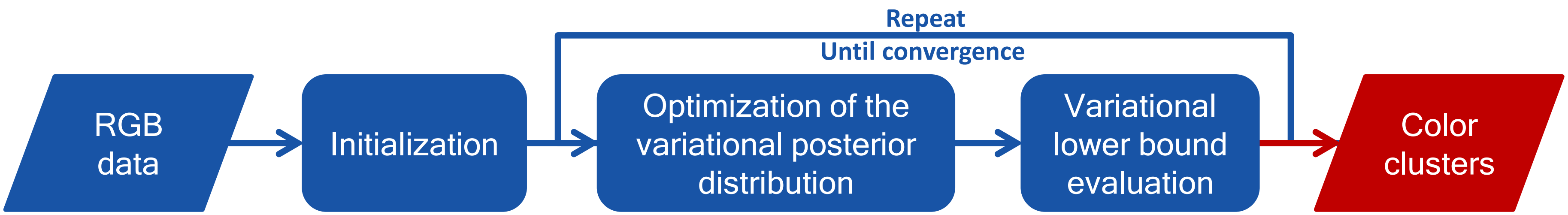
[4] C. M. Bishop, Pattern Recognition and Machine Learning, 1st ed. New York: Springer, 2006.

Gaussian mixture model with variational Bayes inference



[4] C. M. Bishop, Pattern Recognition and Machine Learning, 1st ed. New York: Springer, 2006.

Gaussian mixture model with variational Bayes inference



[4] C. M. Bishop, Pattern Recognition and Machine Learning, 1st ed. New York: Springer, 2006.