IEEE International Conference on Acoustics, Speech, and Signal Processing

# Multiview Depth Map Enhancement by Variational Bayes Inference Estimation of Dirichlet Mixture Models

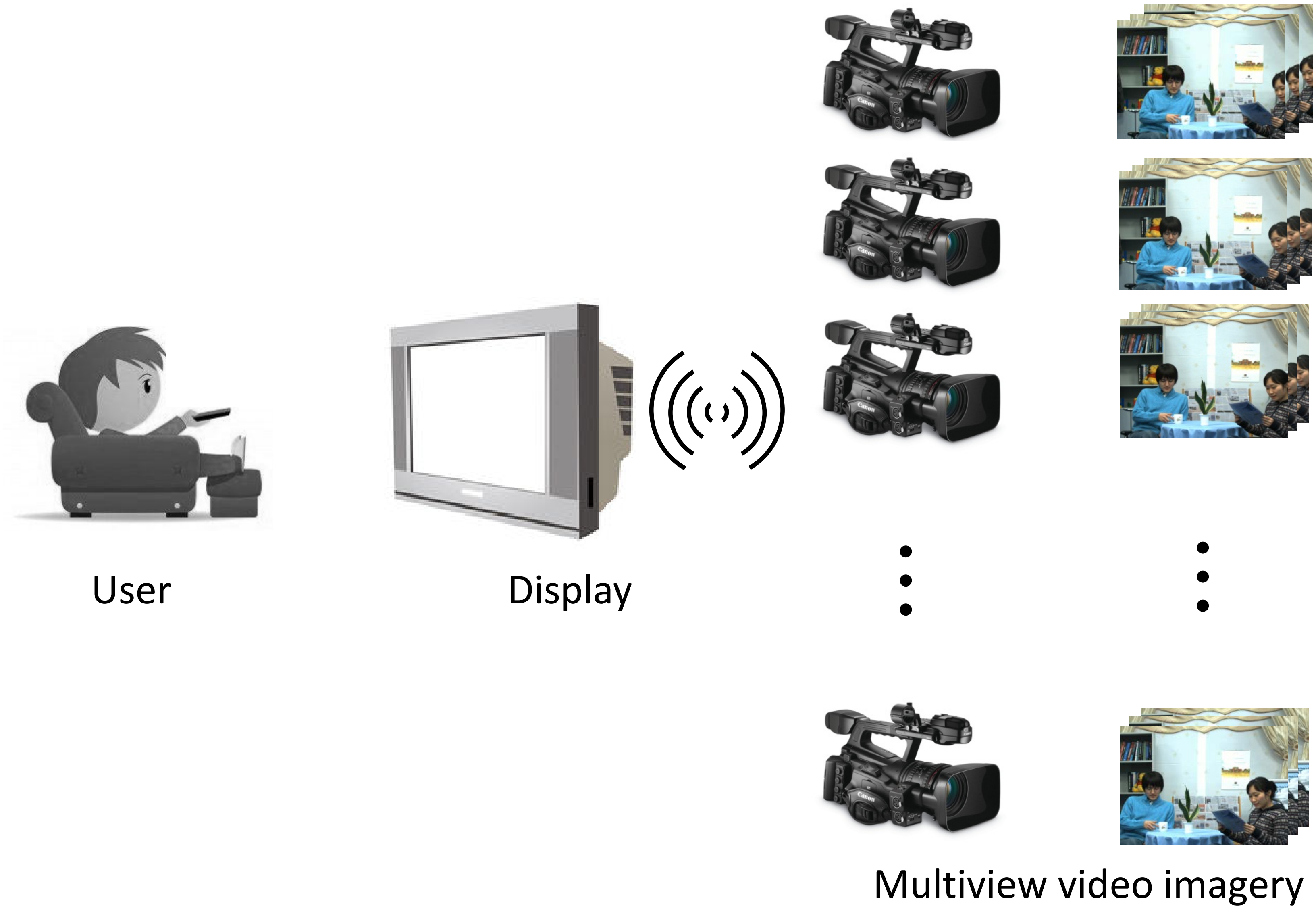
Pravin Kumar Rana, Zhanyu Ma, Jalil Taghia, and Markus Flierl

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

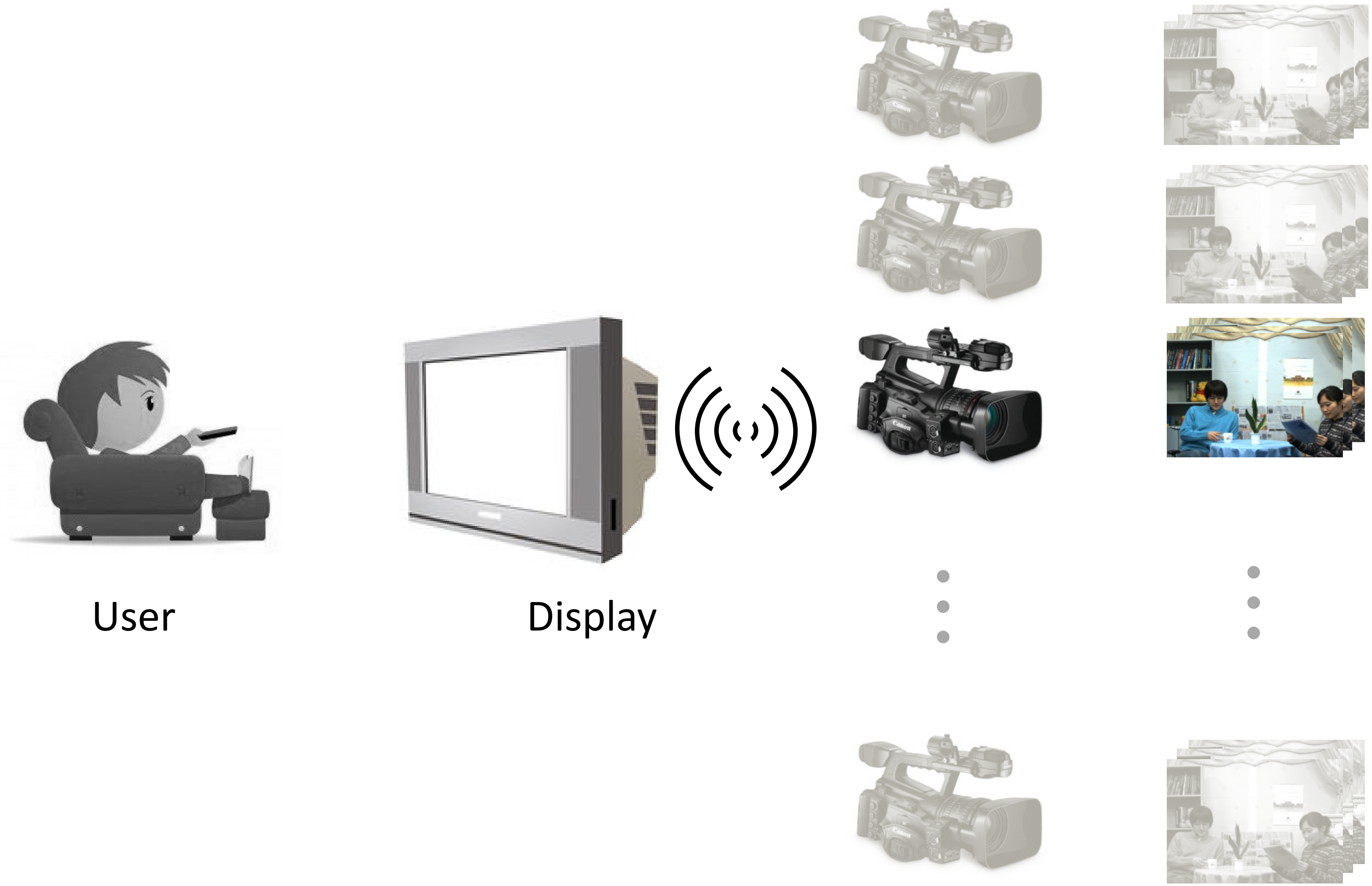
May 31, 2013

# Motivation and background

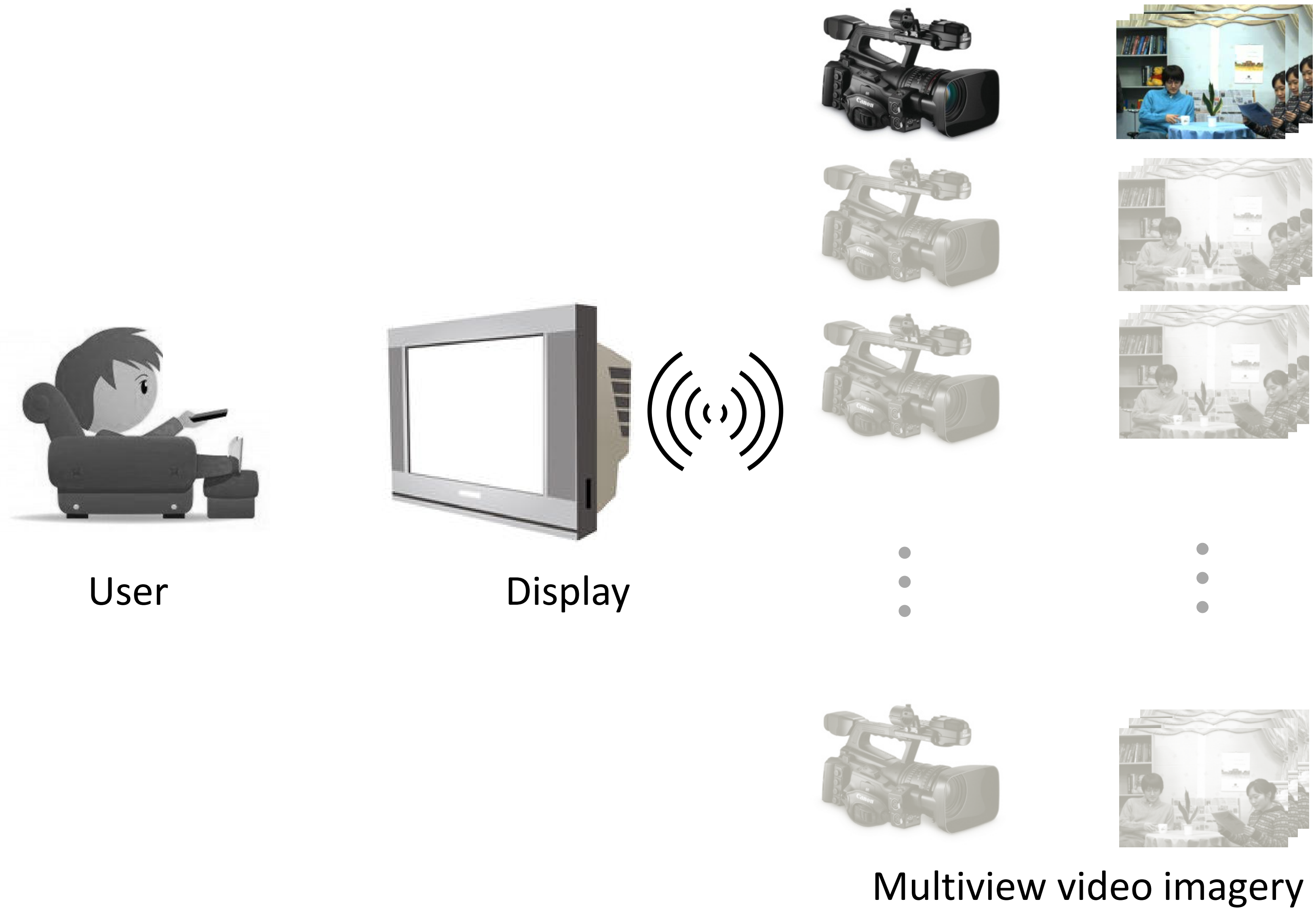
# Free-viewpoint television



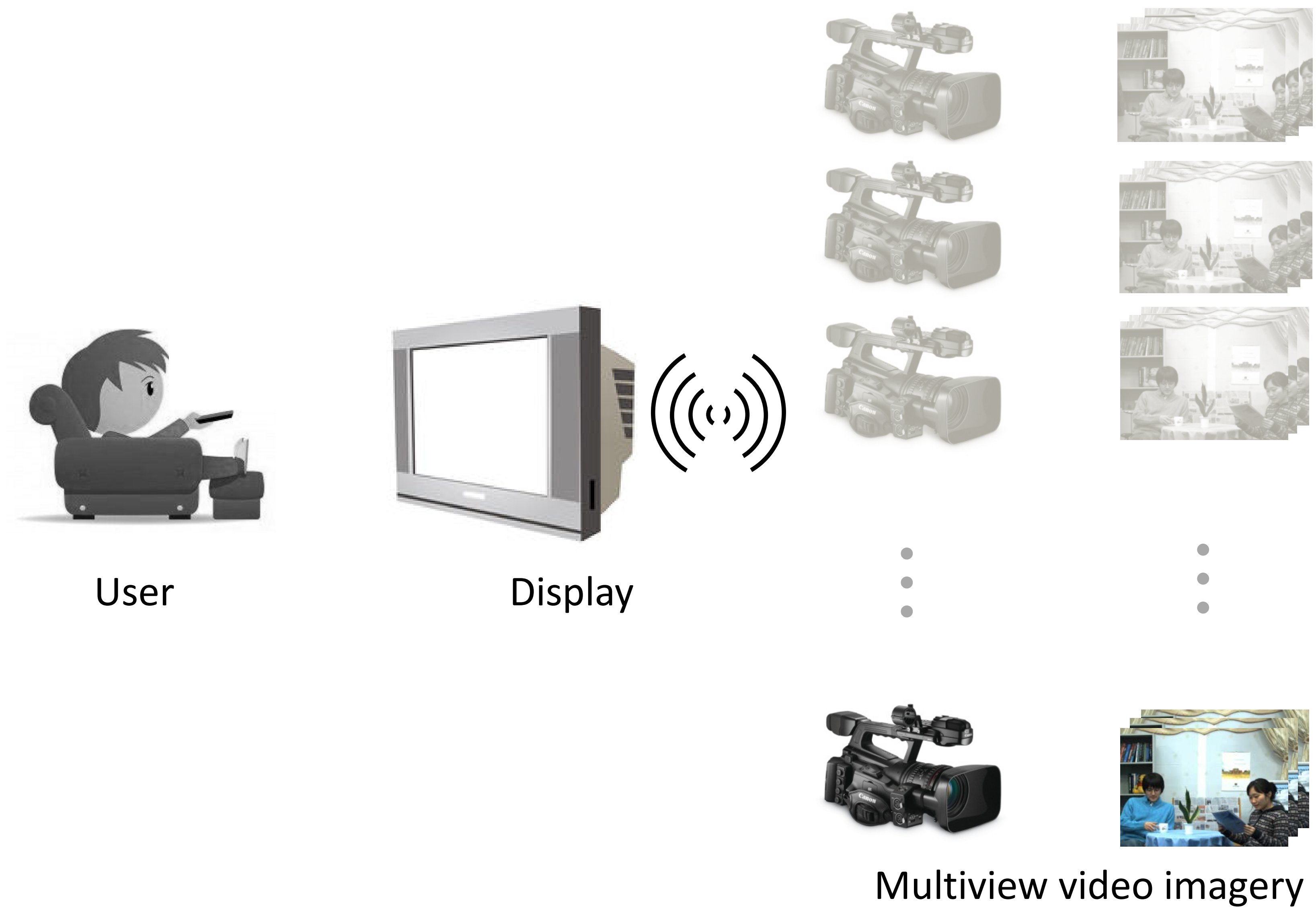
# Free-viewpoint television



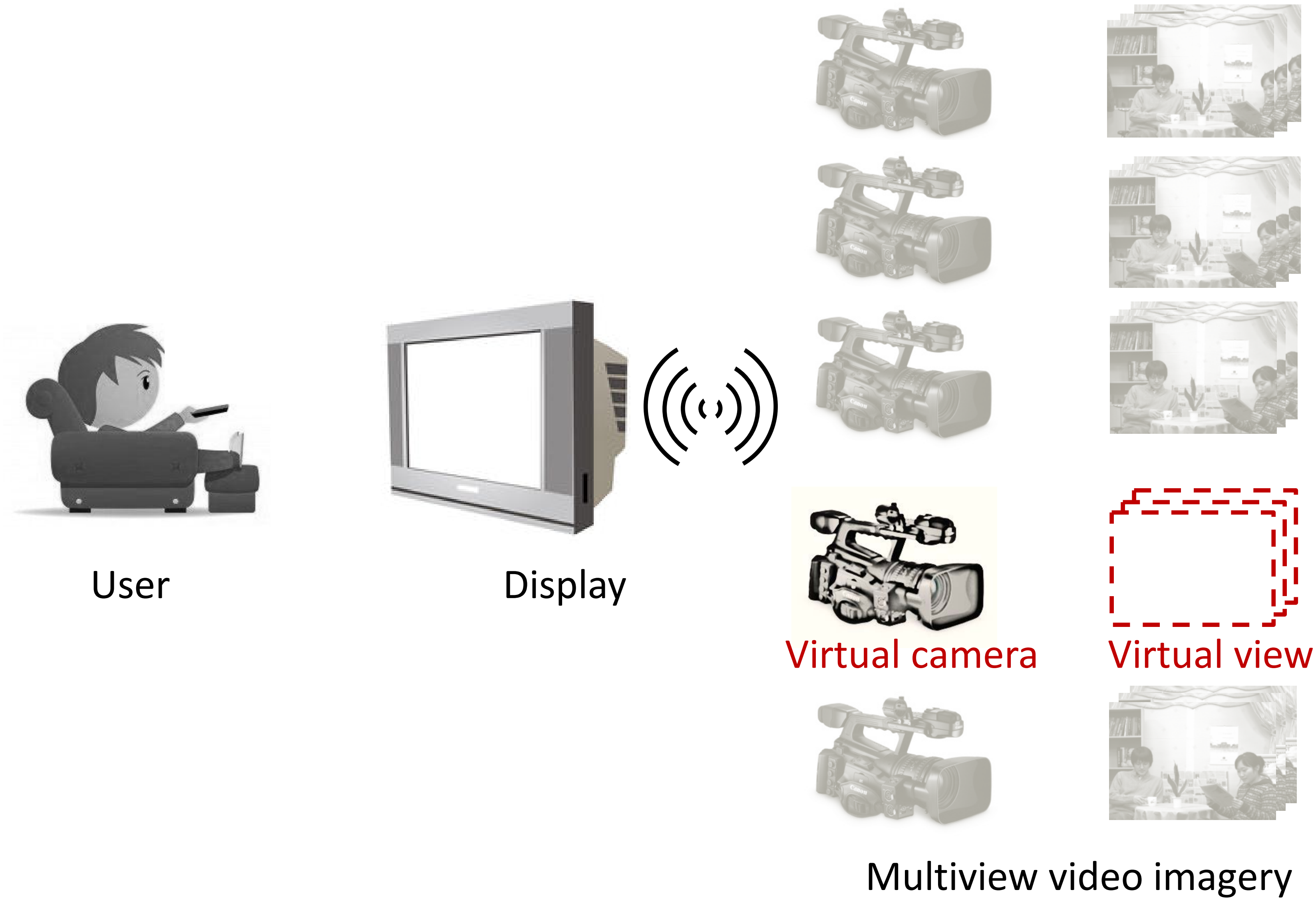
# Free-viewpoint television



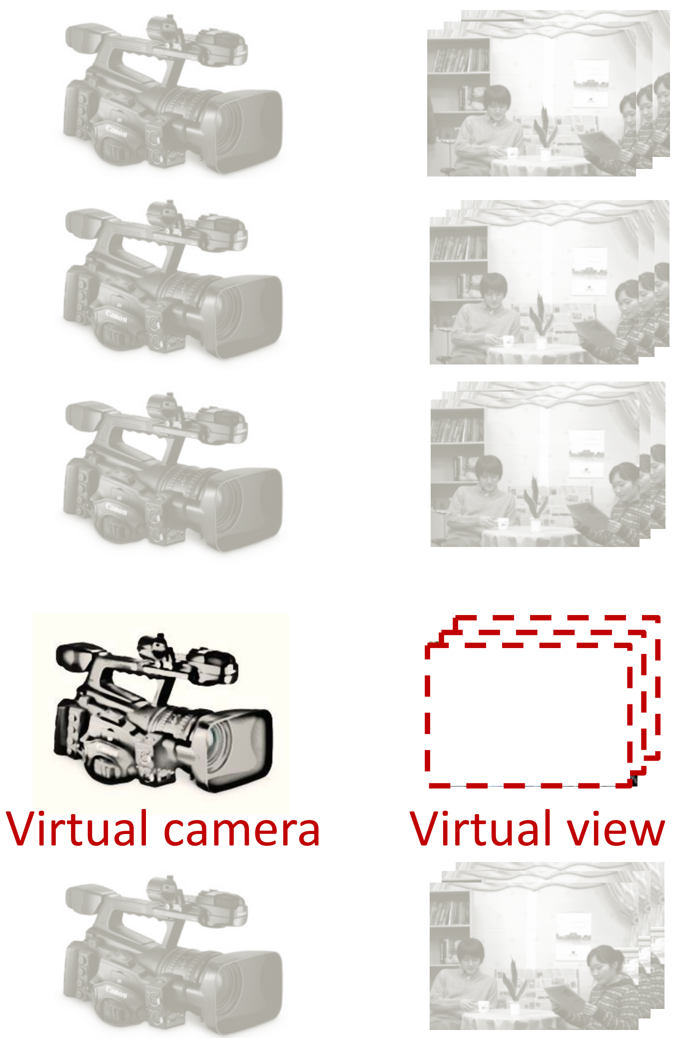
# Free-viewpoint television



# Free-viewpoint television

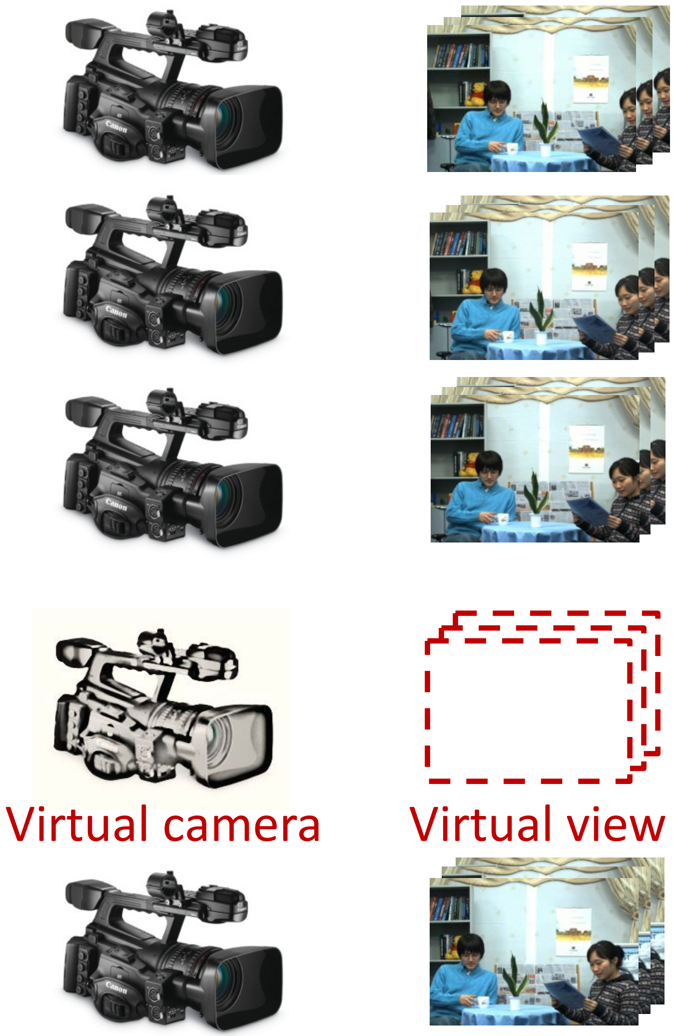


# Depth image based rendering




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

Near

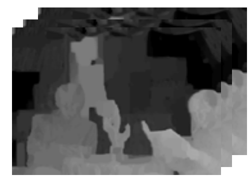
Far

3D warping

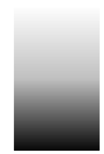
# Depth image based rendering

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

Depth image



Near



Far

3D warping

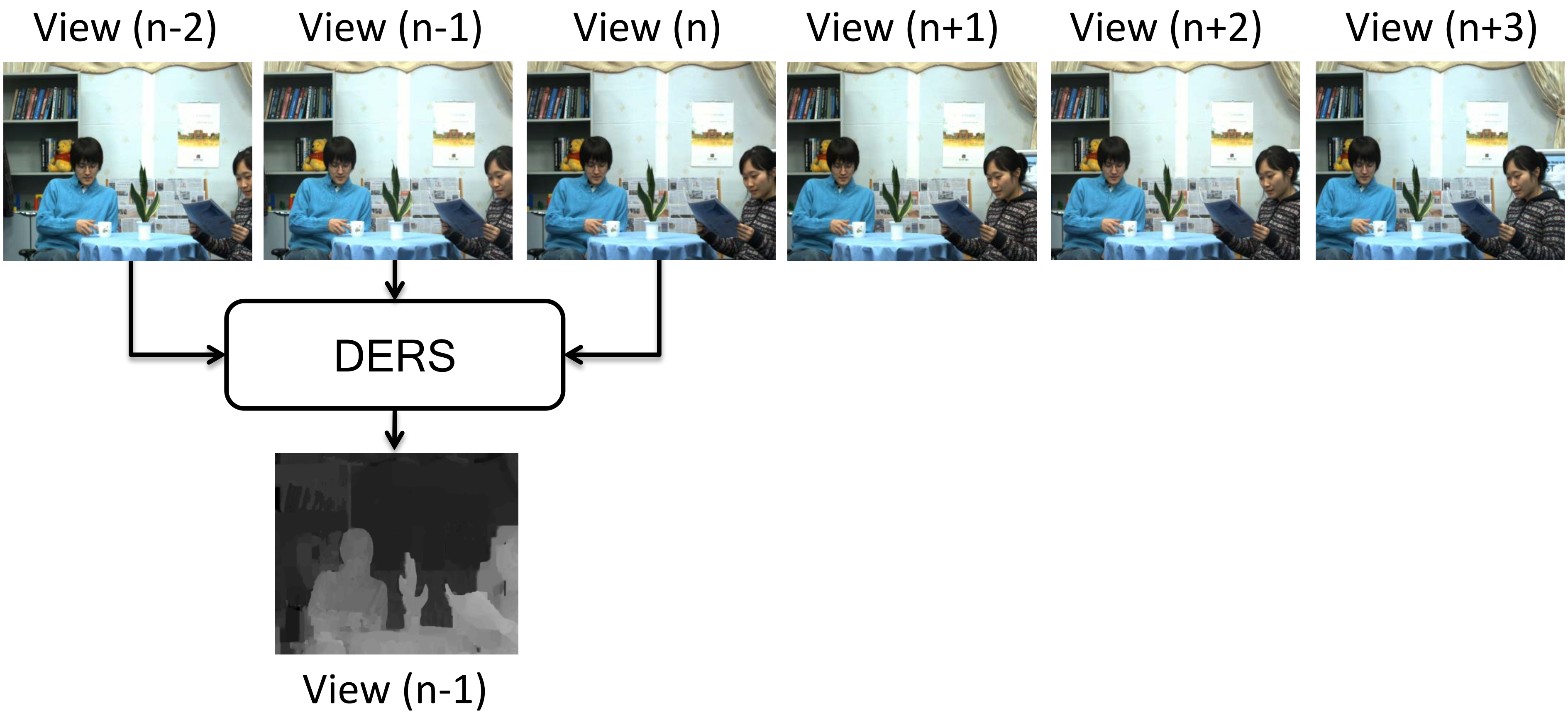
Virtual camera

Virtual view

Multiview video imagery

# Depth estimation

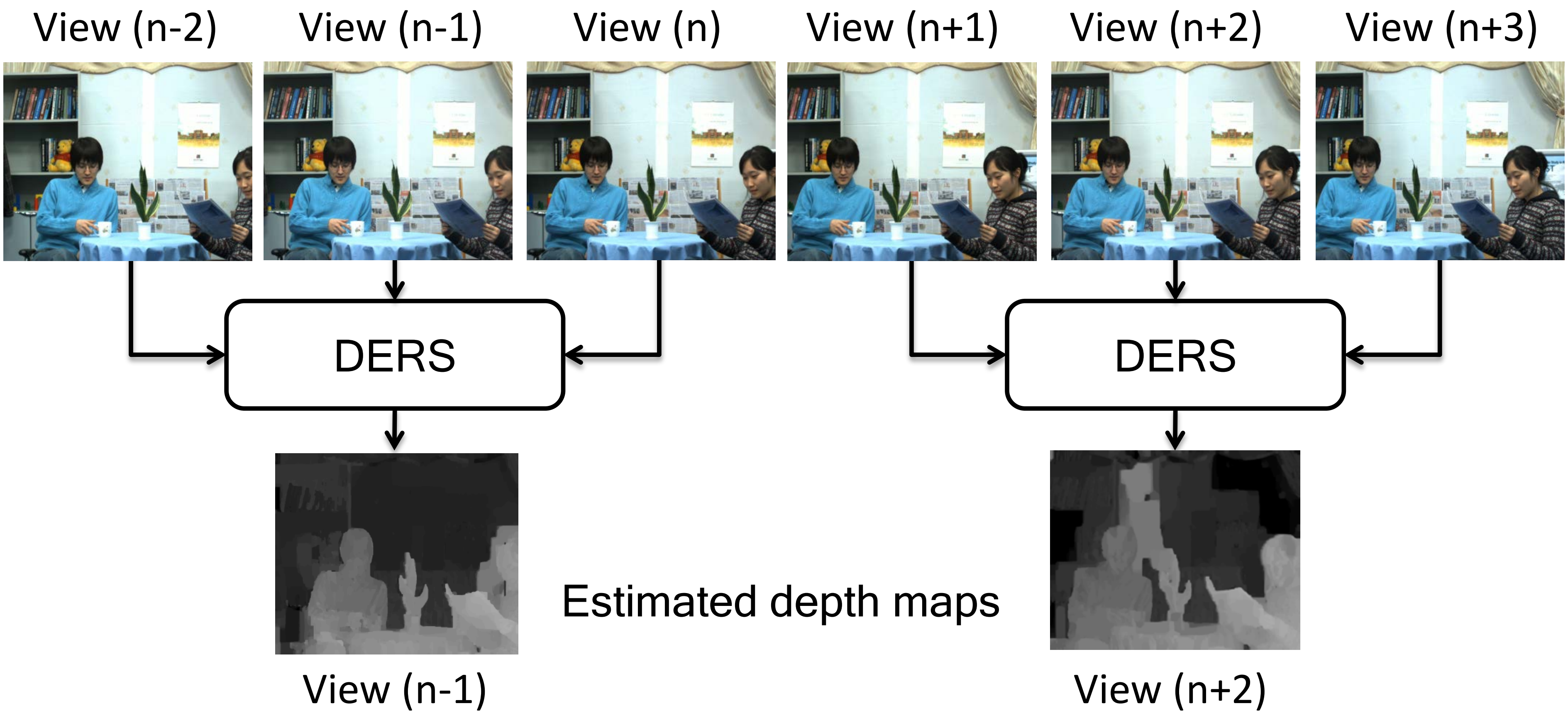
MPEG Depth Estimation Reference Software (DERS)




Note: we assume a 1D-parallel camera arrangement

# Depth estimation

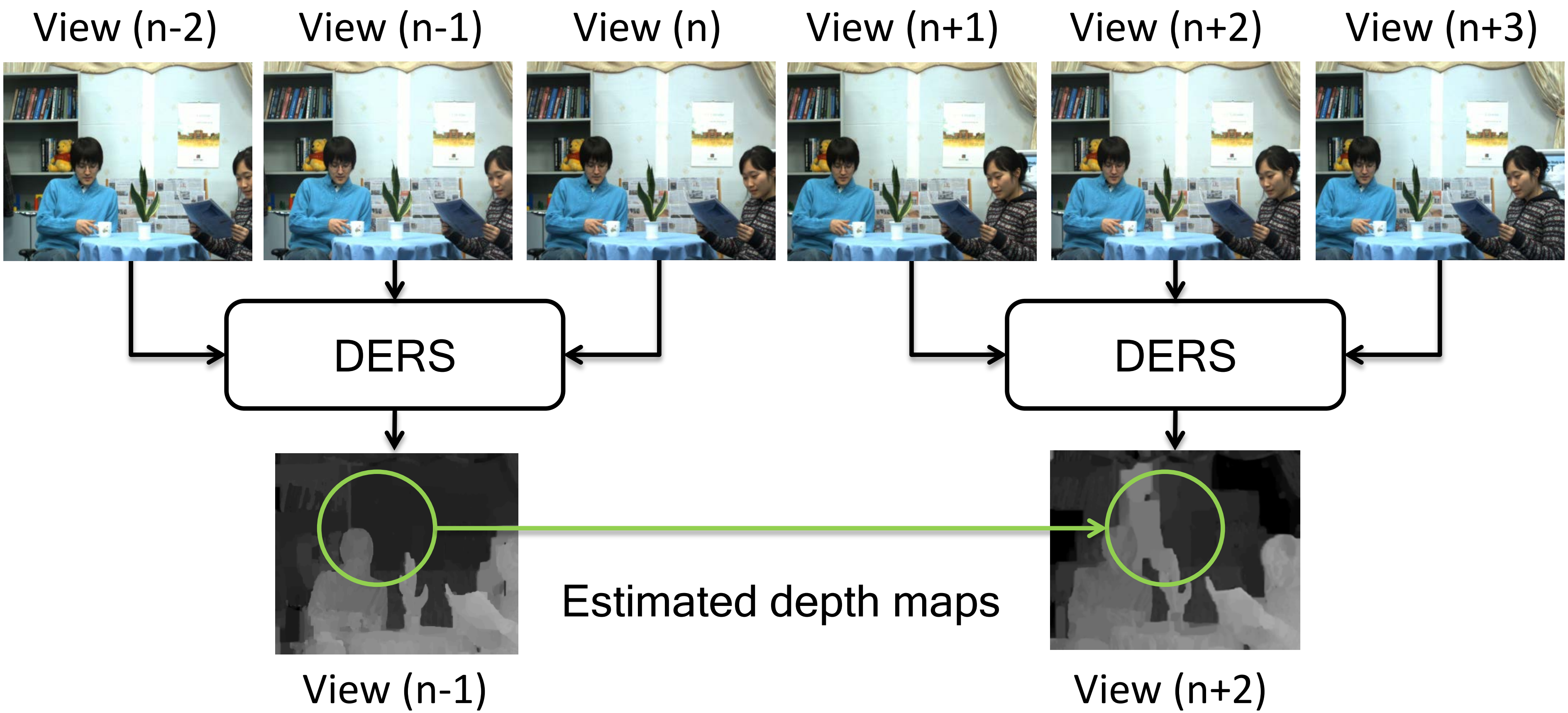
MPEG Depth Estimation Reference Software (DERS)




Note: we assume a 1D-parallel camera arrangement

# Depth estimation

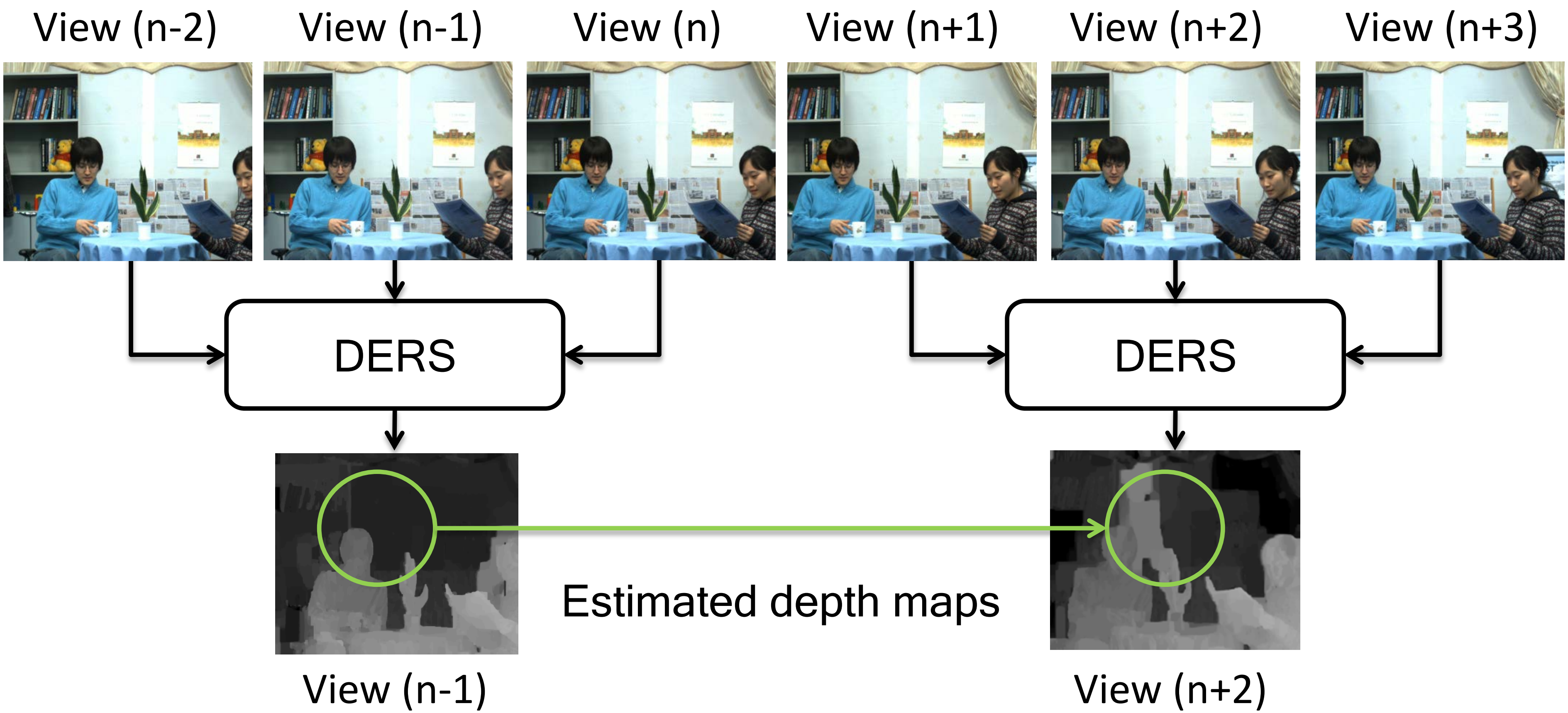
## MPEG Depth Estimation Reference Software (DERS)




Note: we assume a 1D-parallel camera arrangement

# Depth estimation

## MPEG Depth Estimation Reference Software (DERS)




**Problem: Inter-view depth inconsistency**

Note: we assume a 1D-parallel camera arrangement

# Depth enhancement framework

# Overview of our prior work

Multiview imagery

↓

Concatenation of view imagery

↓

Multiview color classification

↓

Multiview depth classification

↓

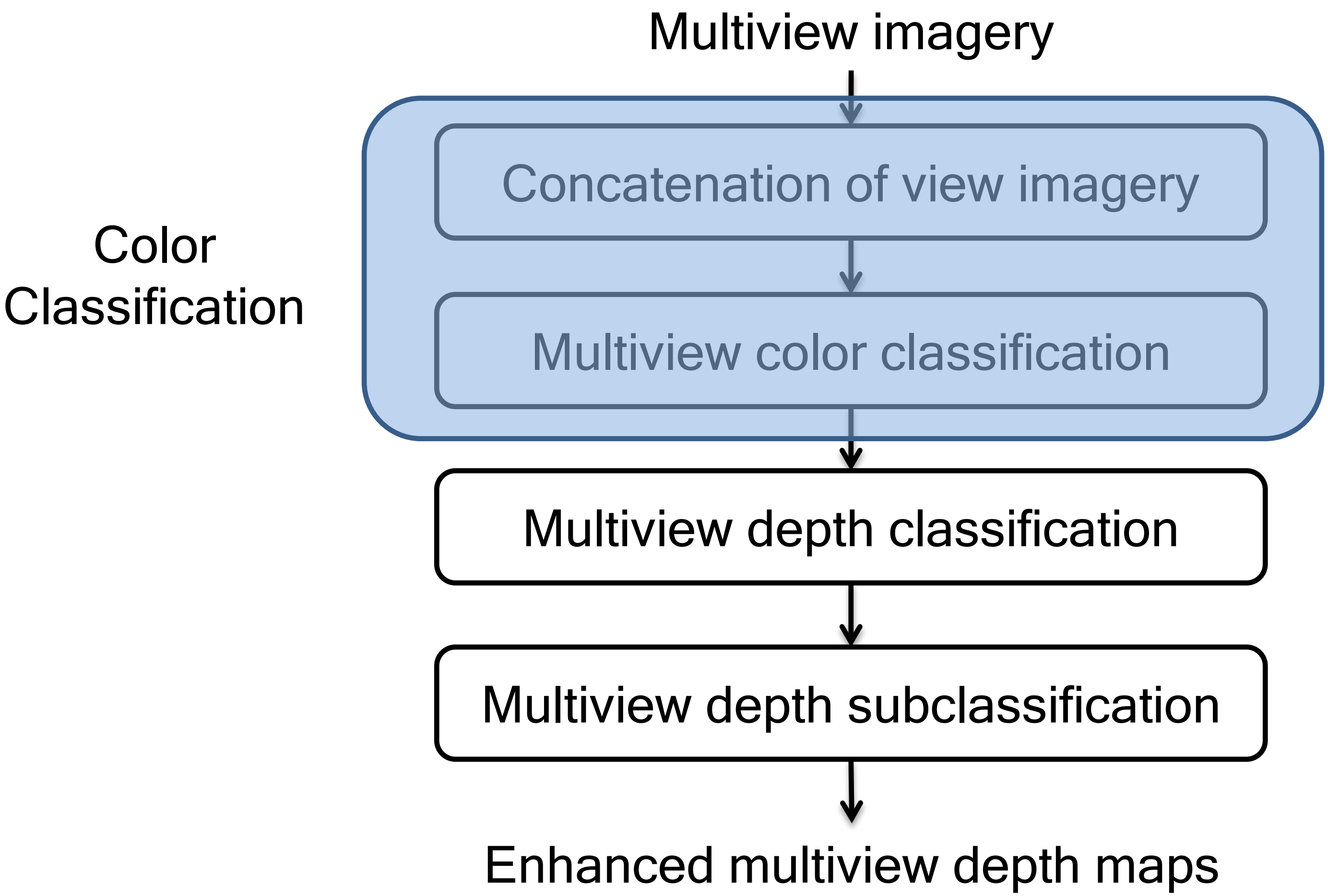
Multiview depth subclassification

↓

Enhanced multiview depth maps

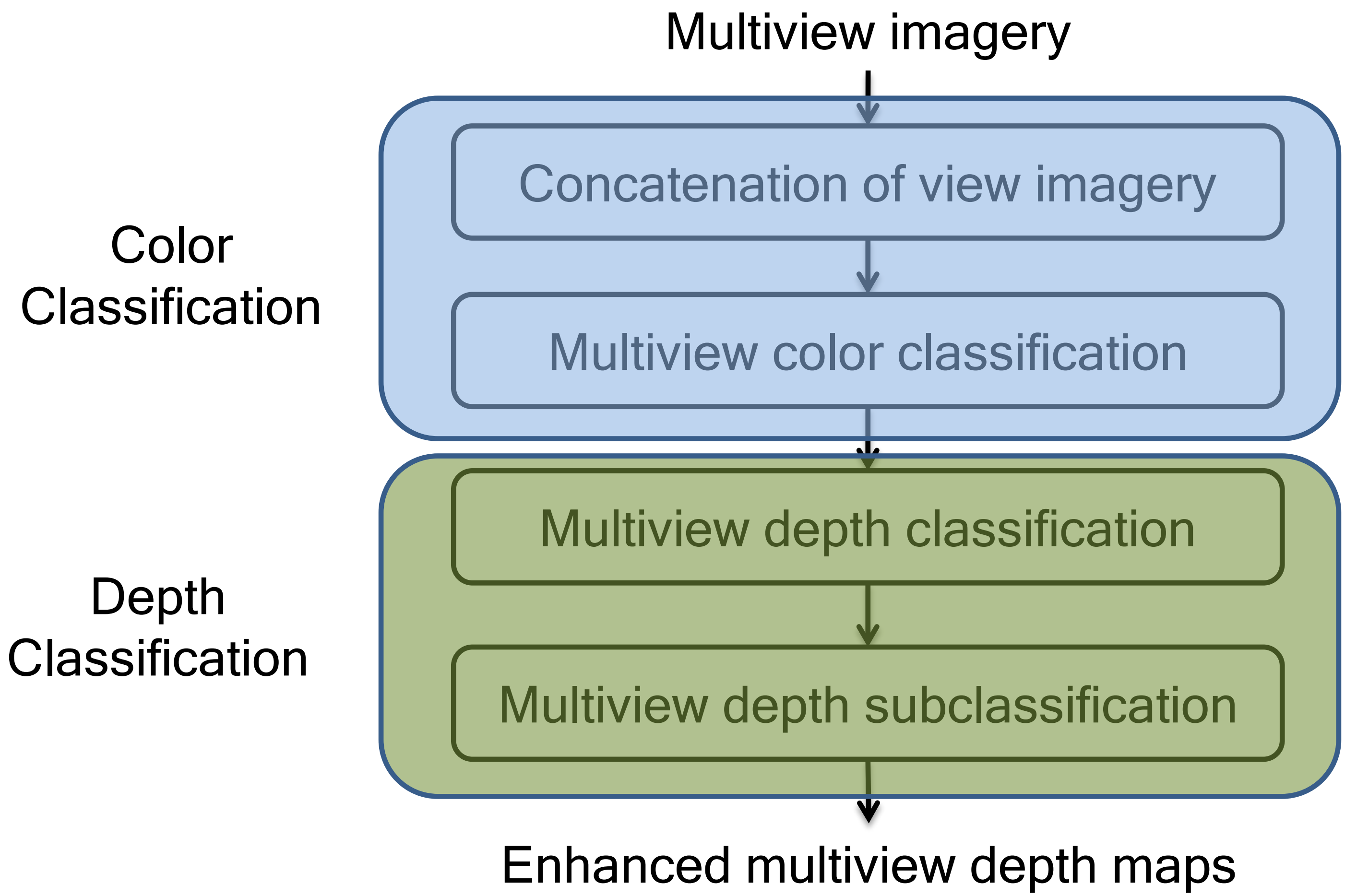
[1] P. K. Rana, J. Taghia, and M. Flierl, "A variational Bayesian inference framework for multiview depth image enhancement," IEEE Int. Symp. Multimedia (ISM), 2012

# Overview of our prior work




[1] P. K. Rana, J. Taghia, and M. Flierl, "A variational Bayesian inference framework for multiview depth image enhancement," IEEE Int. Symp. Multimedia (ISM), 2012

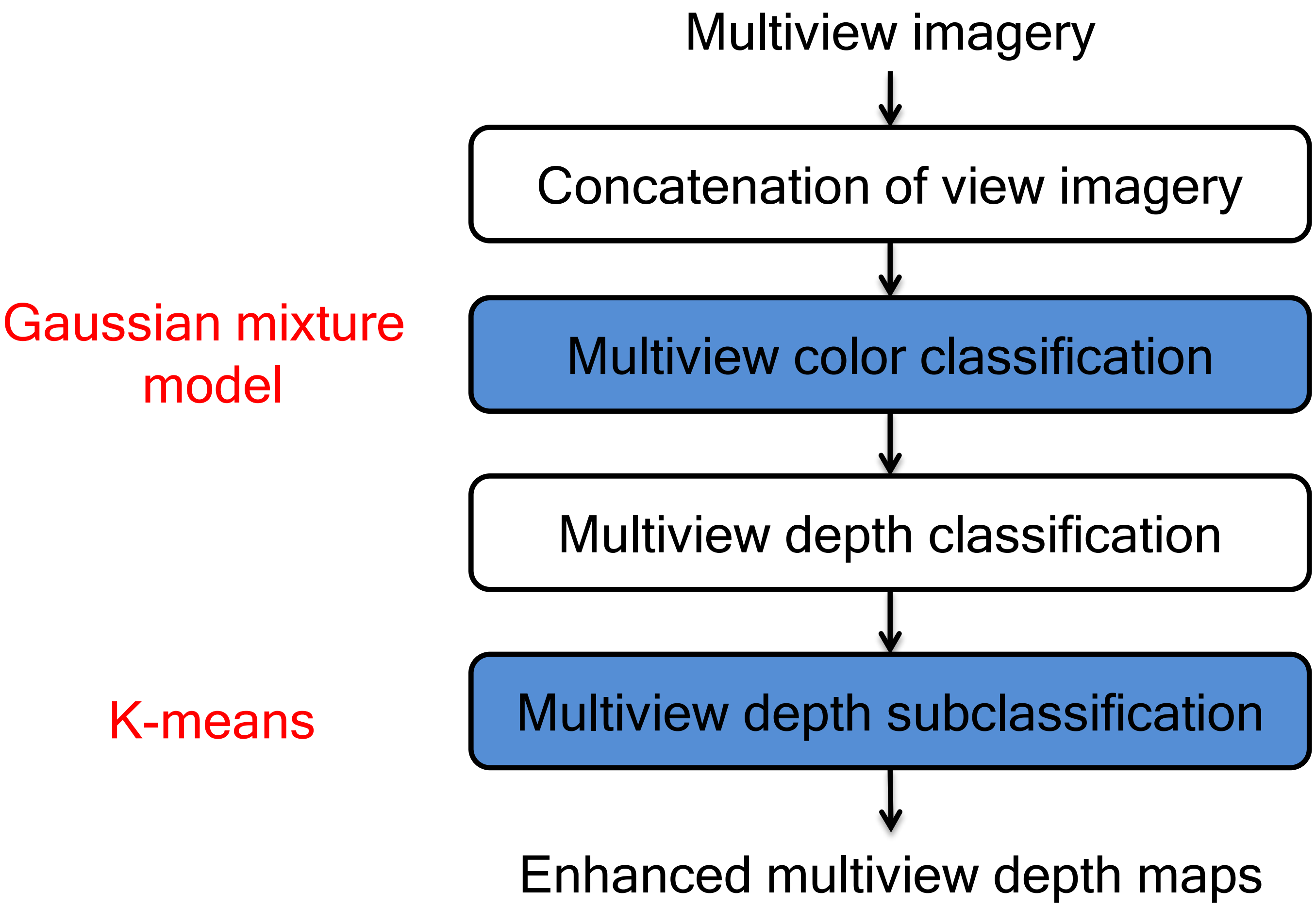
# Overview of our prior work




[1] P. K. Rana, J. Taghia, and M. Flierl, "A variational Bayesian inference framework for multiview depth image enhancement," IEEE Int. Symp. Multimedia (ISM), 2012

# Improved depth enhancement framework

# Improved depth enhancement framework




[1] P. K. Rana, J. Taghia, and M. Flierl, "A variational Bayesian inference framework for multiview depth image enhancement," IEEE Int. Symp. Multimedia (ISM), 2012

# Improved depth enhancement framework

Multiview imagery

↓

Concatenation of view imagery

↓

Gaussian mixture model | **Multiview color classification** | Dirichlet mixture model

↓

Multiview depth classification

↓

K-means | **Multiview depth subclassification** | Mean-shift

↓

Enhanced multiview depth maps

[1] P. K. Rana, J. Taghia, and M. Flierl, "A variational Bayesian inference framework for multiview depth image enhancement," IEEE Int. Symp. Multimedia (ISM), 2012

# Concatenation of view imagery

- Multiview imagery has inherent inter-view similarity
- To have a unique model for multiview imagery
  - The inherent inter-view similarity is exploited by concatenating views from multiple viewpoints

# Multiview color classification

## Color space

View (n) View (n+1)

...  ...

RGB

[1] P. K. Rana, J. Taghia, and M. Flierl, "A variational Bayesian inference framework for multiview depth image enhancement," IEEE Int. Symp. Multimedia (ISM), 2012

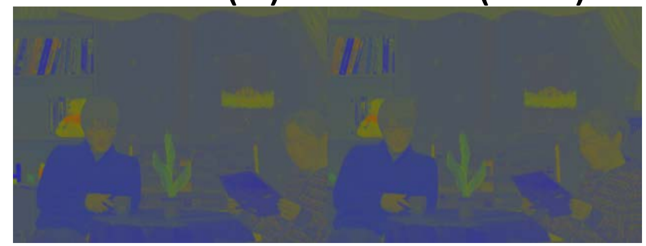
# Multiview color classification

## Color space

View (n) View (n+1)

…  …

RGB

…  …

View (n) View (n+1)

xyz

- Use the chromatic color representation to make the procedure insensitive to the absolut luminance
- The chromaticity of a pixel is described by a vector of three chromaticity coefficients $[x\ y\ z]^T$

$$x+y+z = 1$$

[1] **P. K. Rana, J. Taghia, and M. Flierl, "A variational Bayesian inference framework for multiview depth image enhancement," IEEE Int. Symp. Multimedia (ISM), 2012**

# Multiview color classification

## Why variational Bayes inference (VBI)?

- The goal of classification is to partition an image into regions each of which has a reasonably homogeneous visual appearance
- Usually, clustering algorithm, such as expectation-maximization (EM) suffers from one major drawbacks that the number of clusters has to be known
- Variational Bayes inference automatically select the number of cluster

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

# Multiview color classification

## Why Dirichlet mixture model with variational Bayes inference?

- The vector of image pixels has nonnegative elements and is bounded
  - It can be efficiently modeled by utilizing non-Gaussian distributions [3]
- Based on the pixel vector's properties, assume that the pixel vectors of each cluster are Dirichlet distributed
- Use Dirichlet mixture model (DMM) with VBI to capture the all underlying clusters in multiview imagery
- It reduces complexity

**[3] Z. Ma, P. K. Rana, J. Taghia, M. Flierl, and A. Leijon, "Bayesian estimation of Dirichlet mixture model with variational inference," IEEE Trans. PAMI, submitted, 2013.**

# Multiview color classification

Newspaper

Balloons

Kendo

Input multiview data

# Multiview color classification

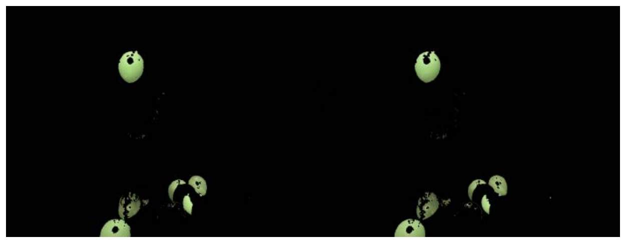
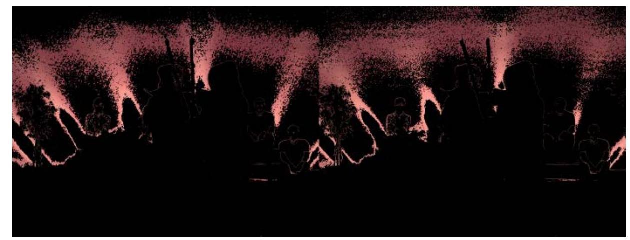
Newspaper

Balloons

Kendo



Input multiview data

Using Dirichlet mixture model with variational Bayes inference

# Multiview color classification

Newspaper | Balloons | Kendo

Input multiview data

Using Dirichlet mixture model with variational Bayes inference

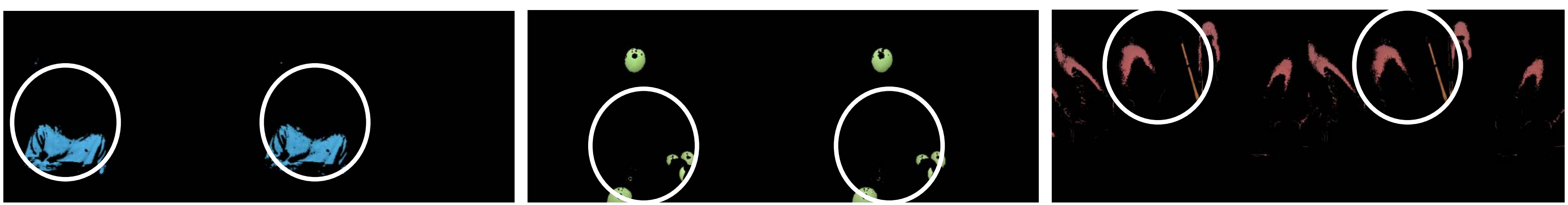
Using Gaussian mixture model with variational Bayes inference

# Multiview color classification

Newspaper | Balloons | Kendo

Input multiview data

Using Dirichlet mixture model with variational Bayes inference



Using Gaussian mixture model with variational Bayes inference

# Multiview depth classification

## Exploiting the per-pixel association between color and depth

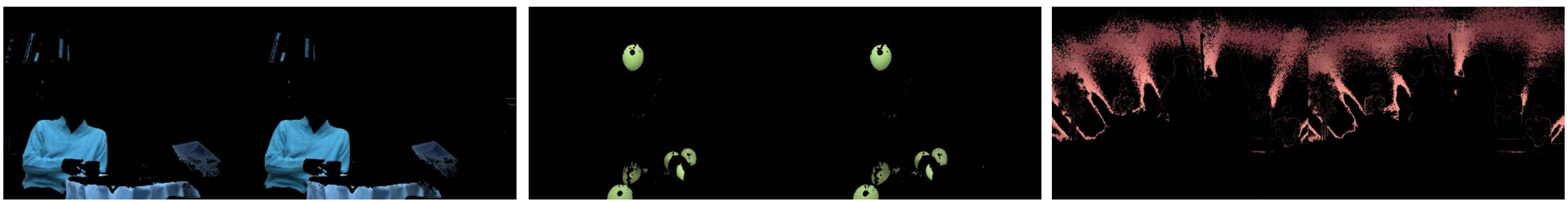
View image Depth image

Concatenated view imagery Concatenated depth imagery

# Multiview depth classification

Newspaper | Balloons | Kendo

Input multiview data



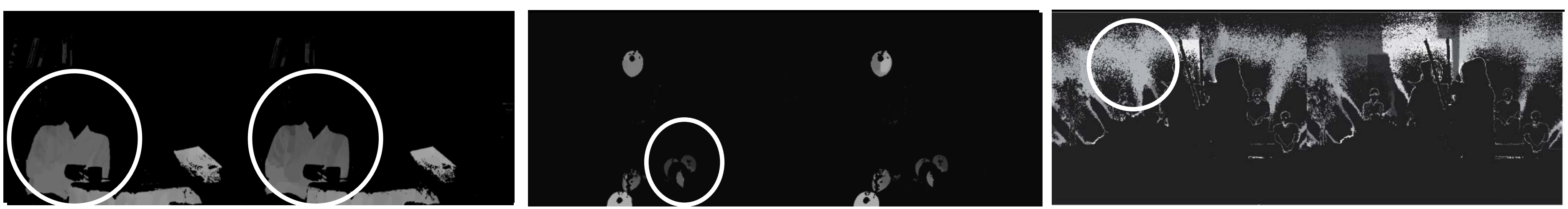
Using Dirichlet mixture model with variational Bayes inference in xyz space

# Multiview depth classification

Newspaper Balloons Kendo

Input multiview data

Using Dirichlet mixture model with variational Bayes inference in xyz space



# Multiview depth subclassification

## 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 subclassification

## Means-shift clustering

- A nonparametric clustering  technique
- Does not require prior knowledge of the number of clusters
- Does not constrain the shape of the clusters
- Assigns the mean to depth pixels irrespective of the originating viewpoints
- Bayesian approaches imply higher computational complexity

# Experimental results

# 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)

# Complexity

| Multiview data set | Initial number of mixture components | Active number of mixture components (after convergence) | |
|---|---|---|---|
| | | VBI-GMM | VBI-DMM |
| Lovebird1 | 100 | 31 | 24 |
| Kendo | 100 | 34 | 15 |

# 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

Enhanced depth map

Left

Reference view

Enhanced depth map

Right

Reference view

# 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

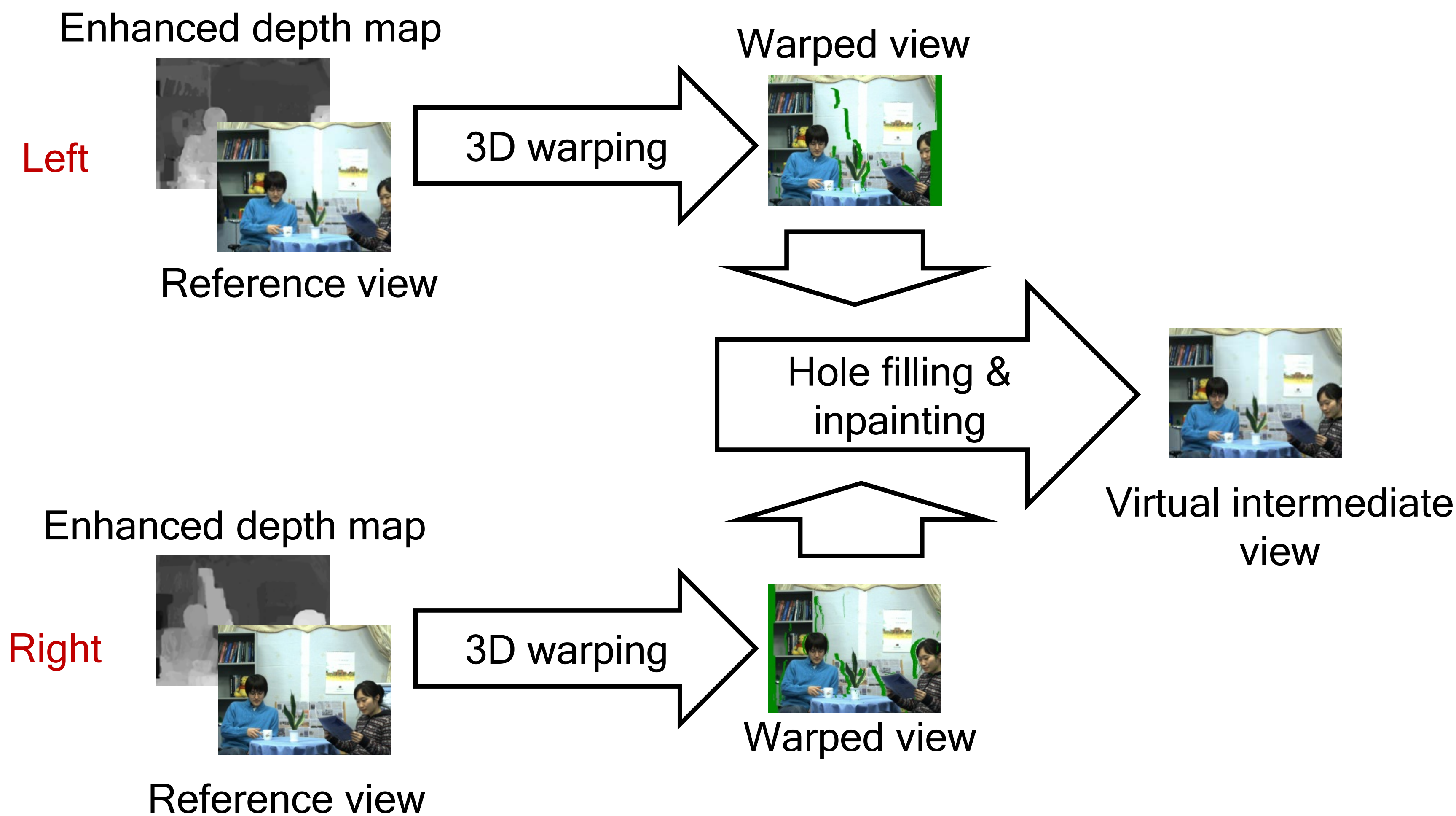
Reference view

3D warping

Warped view

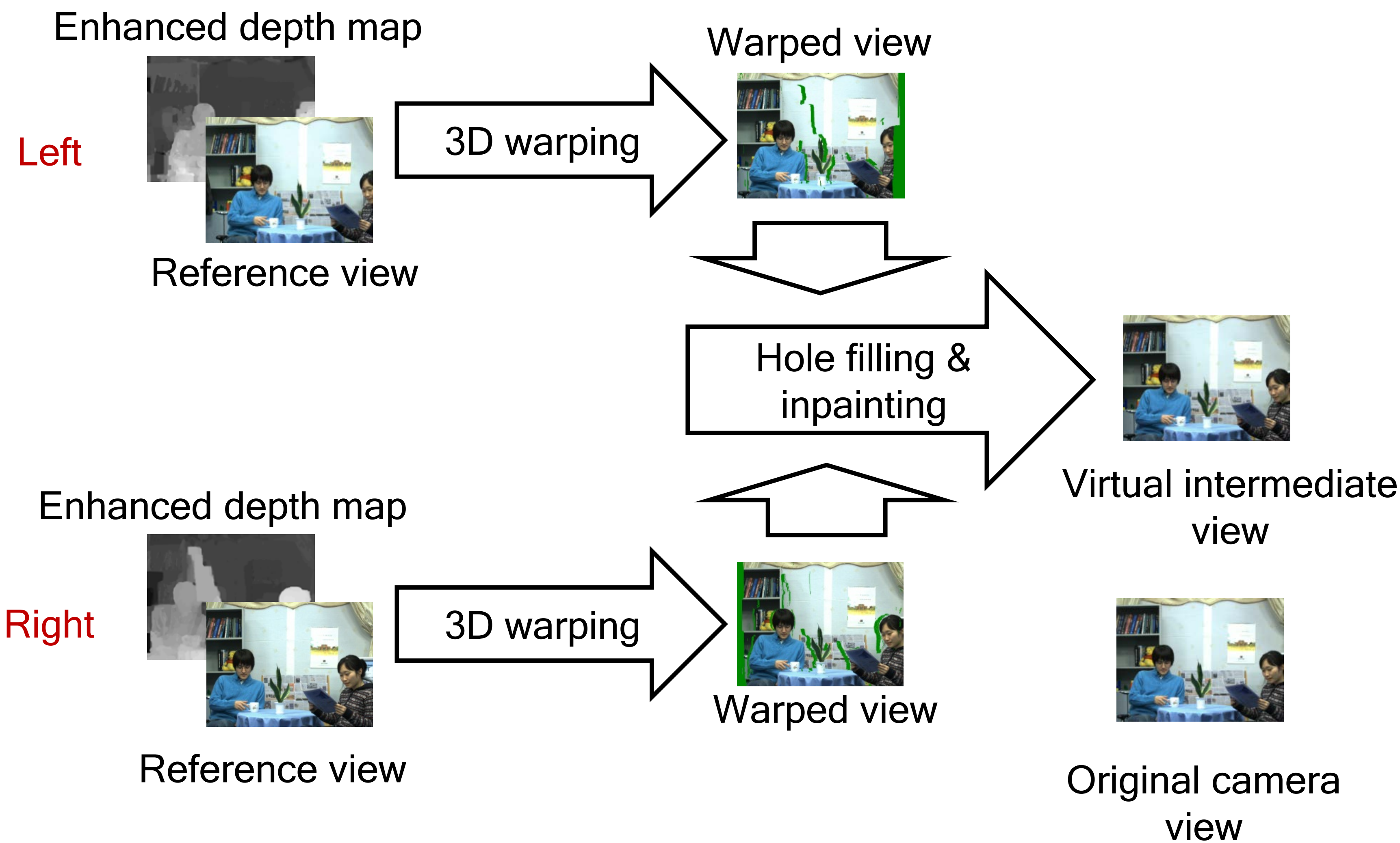
# Depth image based rendering

## MPEG View Synthesis Reference Software (VSRS) 3.5



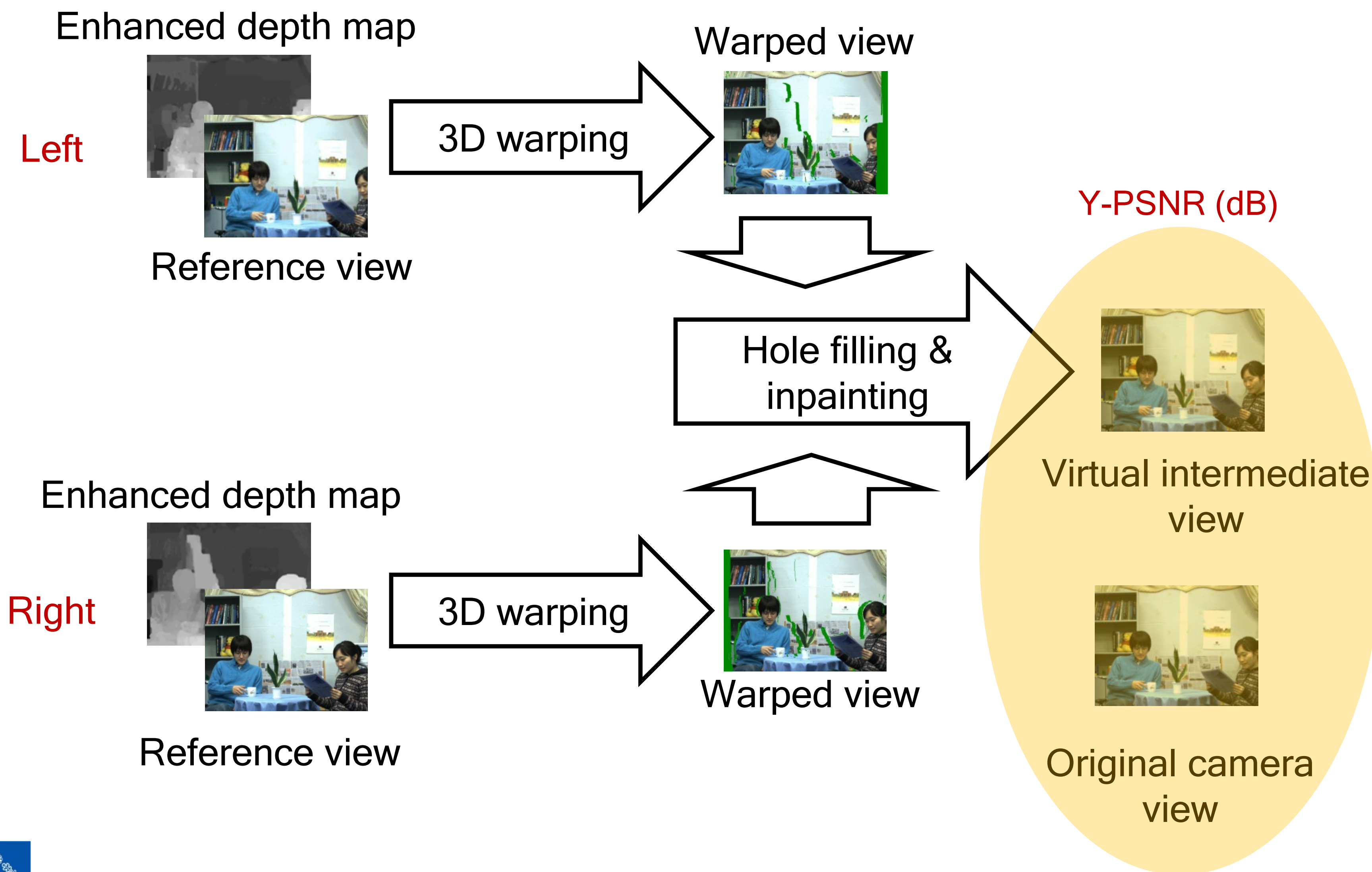
# Depth image based rendering

## MPEG View Synthesis Reference Software (VSRS) 3.5



# Depth image based rendering

## MPEG View Synthesis Reference Software (VSRS) 3.5



# Objective results

| Test sequence | Input view pair | Virtual view | Y-PSNR [dB] | | |
|---|---|---|---|---|---|
| | | | With MPEG depth maps | With VBIGMM + K-Means depth maps | With VBIDMM + Mean-shift depth maps |
| Newspaper | (4, 6) | 5 | 32.00 | 32.10 | 32.11 |
| Kendo | (3, 5) | 4 | 36.54 | 36.72 | 39.35 |
| Lovebird1 | (6, 8) | 7 | 28.50 | 28.68 | 29.04 |
| Balloons | (3, 5) | 4 | 35.69 | 35.93 | 36.02 |
| Poznan Street | (3, 5) | 4 | 35.56 | 35.58 | 35.72 |

- K-means sub-clustering
  - Number of cluster : 12

# Objective results

| Test sequence | Input view pair | Virtual view | Y-PSNR [dB] | | |
|---|---|---|---|---|---|
| | | | With MPEG depth maps | With VBIGMM + K-Means depth maps | With VBIDMM + Mean-shift depth maps |
| Newspaper | (4, 6) | 5 | 32.00 | 32.10 | 32.11 |
| Kendo | (3, 5) | 4 | 36.54 | 36.72 | 39.35 |
| Lovebird1 | (6, 8) | 7 | 28.50 | 28.68 | 29.04 |
| Balloons | (3, 5) | 4 | 35.69 | 35.93 | 36.02 |
| Poznan Street | (3, 5) | 4 | 35.56 | 35.58 | 35.72 |

- K-means sub-clustering
  - Number of cluster : 12

# Subjective results

## Test sequence: Kendo

With MPEG depth map

With VBDMM Mean-shift depth map

# Subjective results

## Test sequence: Kendo

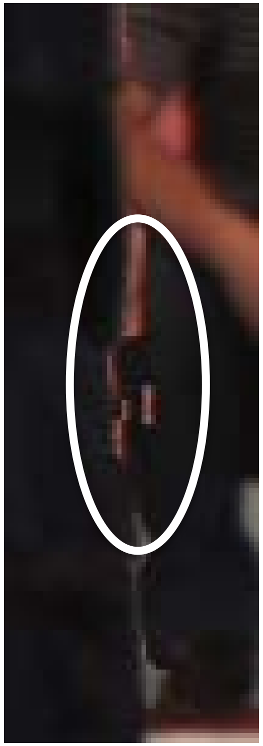
With MPEG depth map

With VBDMM Mean-shift depth map

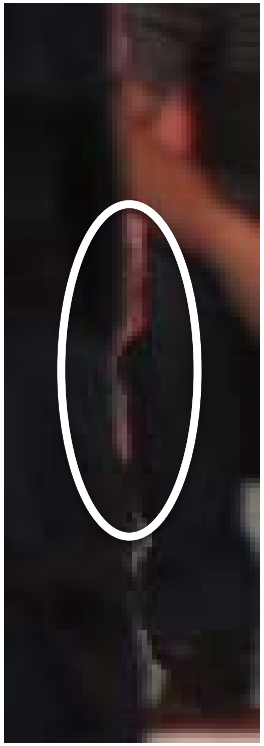
# Subjective results

Test sequence: Kendo

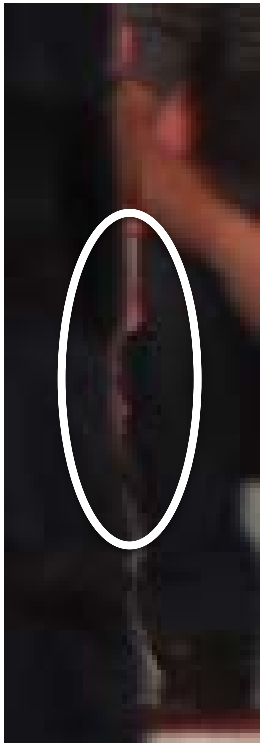
Original



With MPEG depth maps



With VBGMM + K-Means depth maps

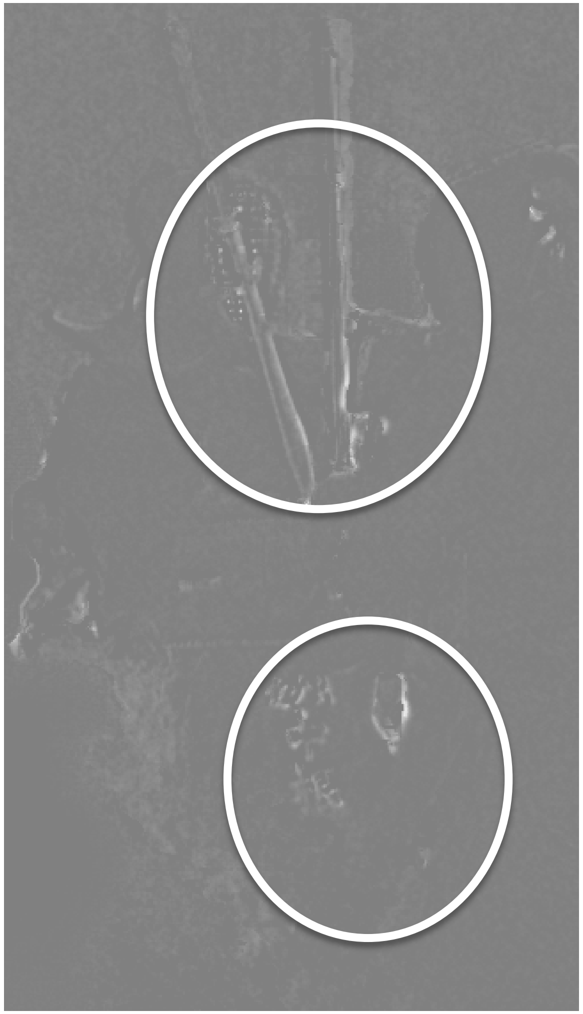


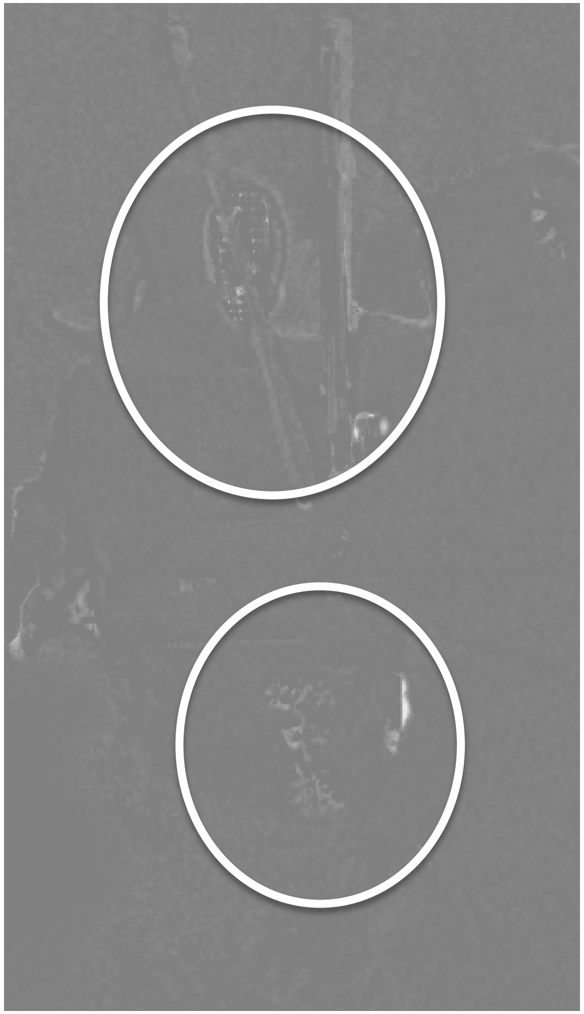
With VBDMM + Mean-Shift depth maps

# Subjective results

Test sequence: Kendo

With VBGMM
K-Means
depth maps





With VBDMM
Mean-Shift
depth maps

# Subjective results

Test sequence: Lovebird 1

With MPEG depth map

With VBDMM Mean-shift depth map

# Subjective results

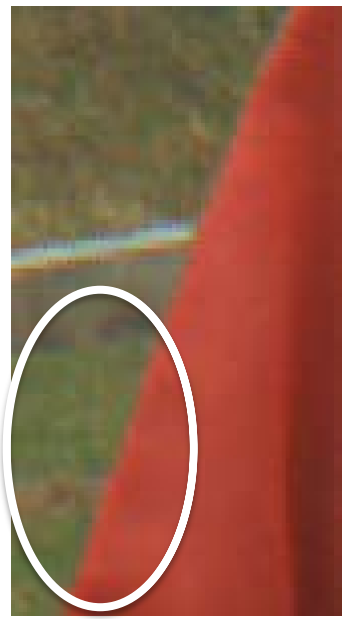
Test sequence: Lovebird 1

With MPEG depth map

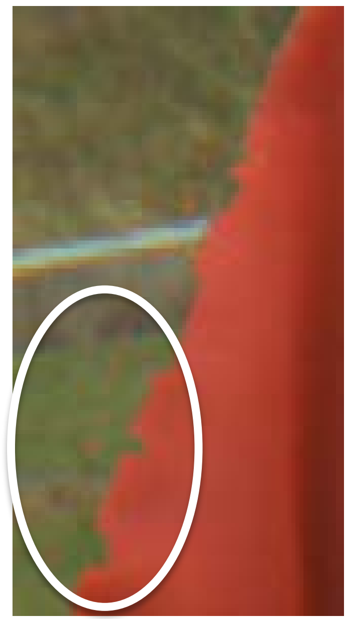
With VBDMM Mean-shift depth map

# Subjective results

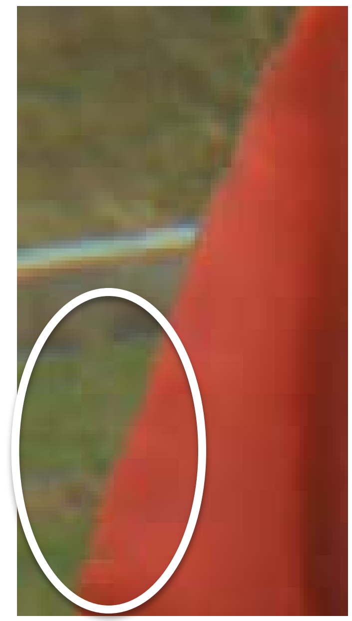
Test sequence: Lovebird 1



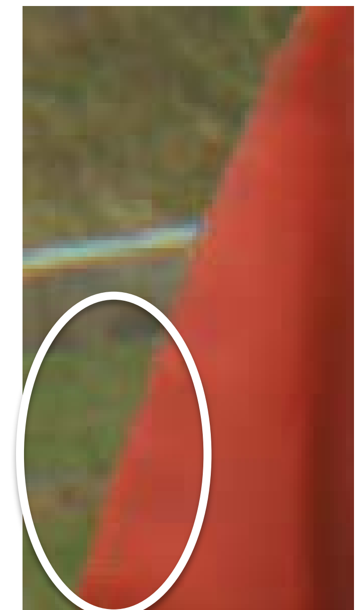
Original



With MPEG depth maps



With VBGMM K-Means depth maps



With VBDMM Mean-Shift depth maps

# Subjective results

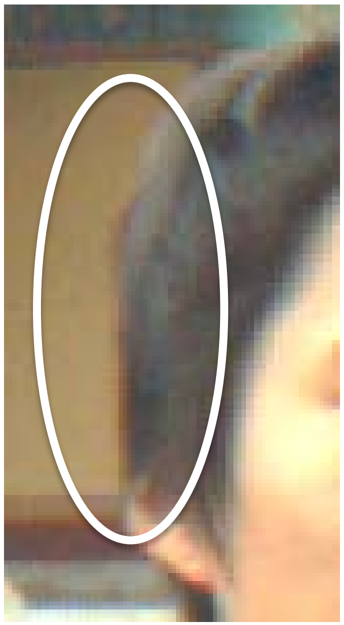
Test sequence: Lovebird 1

With MPEG depth map

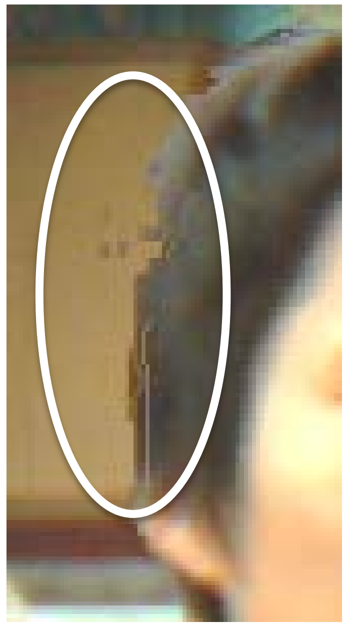
With VBDMM Mean-shift depth map

# Subjective results

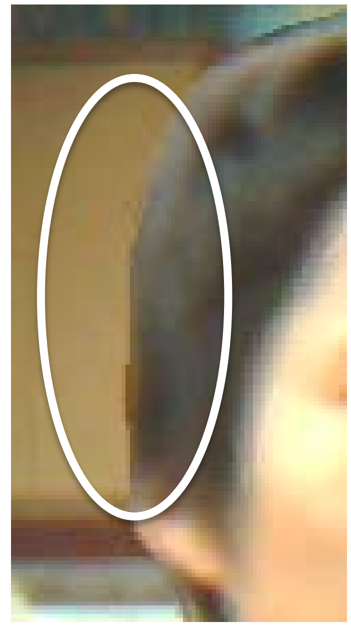
Test sequence: Lovebird 1



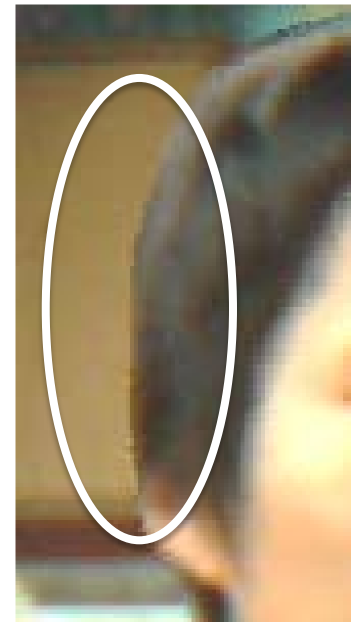
Original



With MPEG
depth maps



With VBGMM
K-Means
depth maps



With VBDMM
Mean-Shift
depth maps

# Conclusions

- The inter-view depth consistency and hence, the free-viewpoint experience improve
- The per-pixel association between depth and color is exploited by classification
- Depth subclassification improves depth maps and hence, view rendering quality
- Both objective and subjective results improve

# Thank you