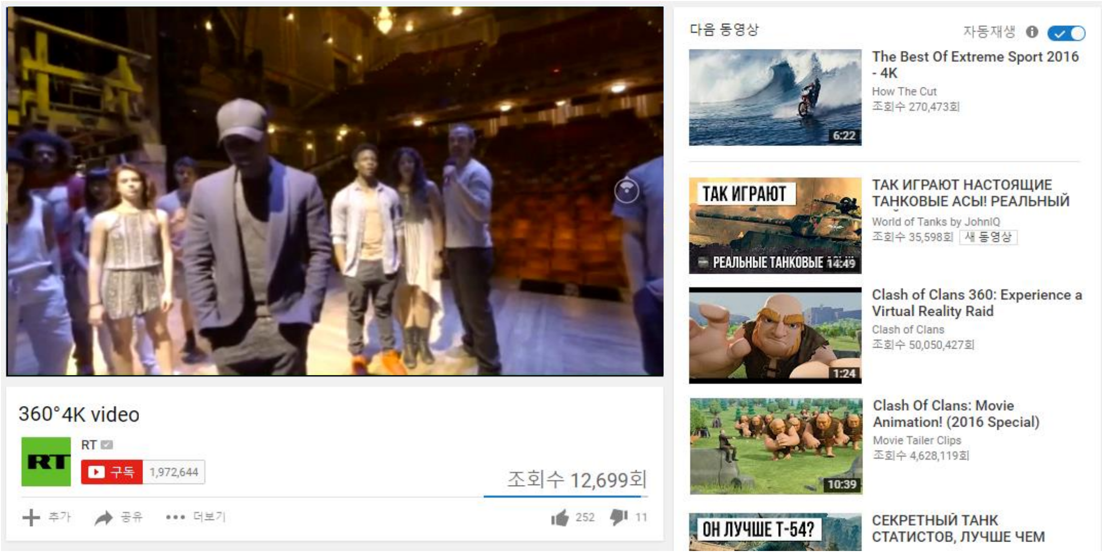
Automatic Viewpoint Selection on 360° Videos

Presenter: Dae-Yong

Contents

Motivation

- **Emergence of 360 contents; Virtual Reality (VR)**
  - Contents produced by 360° images
    - **360° image** is the image that captures **all scenes around the camera without any blind spot**.
  - Recently, numerous global companies launched/provided 360° contents.
    - Facebook and YouTube provide **360° video streaming services** on their websites.
    - Samsung and LG launched **portable 360° cameras** that can easily create 360° images/videos.



< 360° video player on YouTube[1] >

Samsung Gear 360 LG 360 Cam

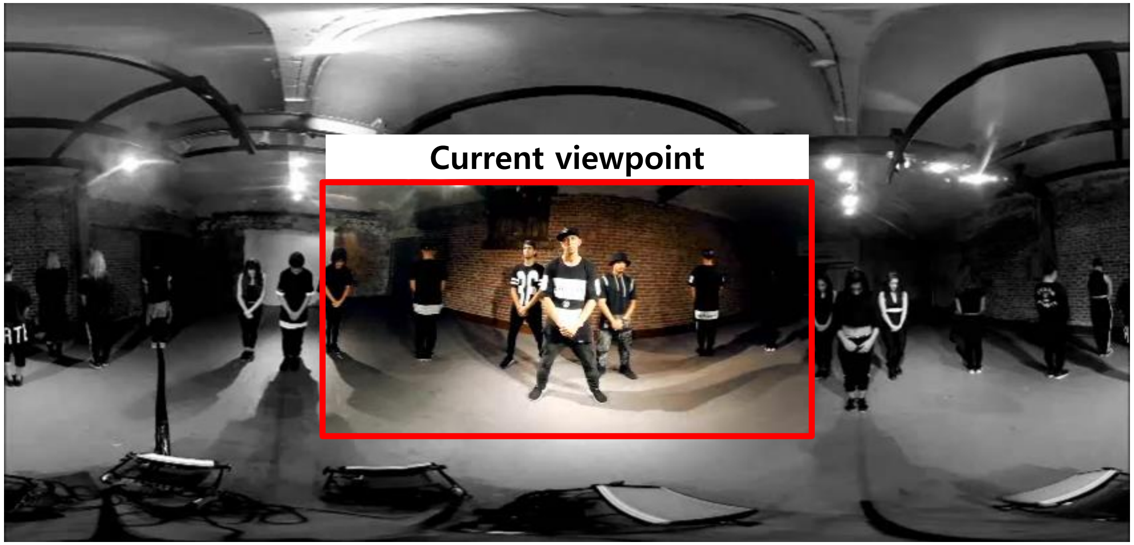
< Portable 360° cameras >

[1] YouTube : https://www.youtube.com/

Motivation

Important issue on watching 360° contents

- Manual operation for viewpoint selection
  - Users **can not catch meaningful events** occurred outside of the current viewpoint.



< Limitation of manual selection of viewpoint>

➡ Generating **perspective videos** from 360° videos via **automatic viewpoint selection**.

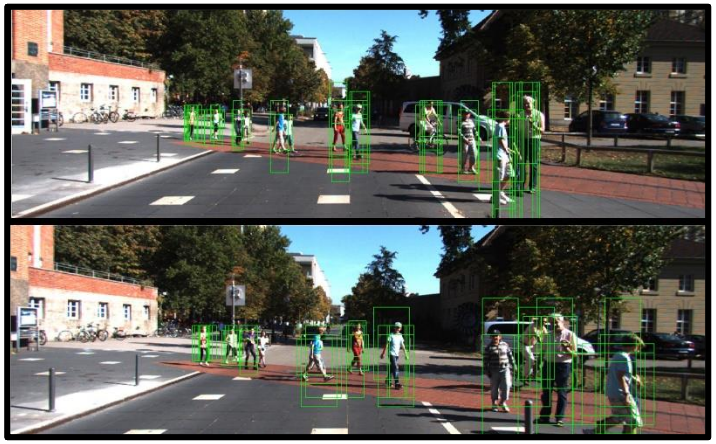
Automatic Viewpoint Selection

Key Assumption

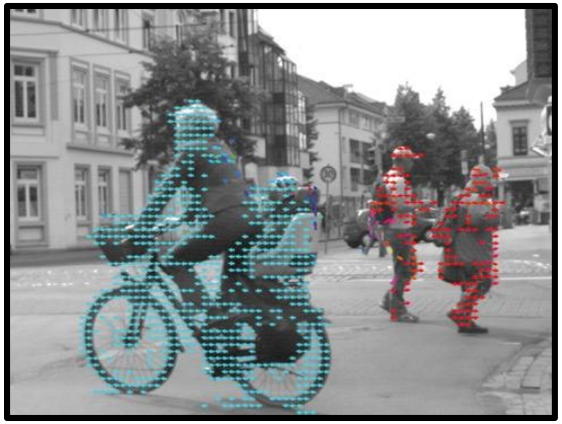
- Better viewpoint contains **MORE SALIENT CONTENTS** on the projected image.

➡ **Analyze visual saliency**, then **select** a **viewpoint**

- Consideration of both an appearance and a motion of the objects
  - motion → Optical flow [5]
  - appearance → Objectness-based object proposal [2,3,4]



< Object Proposal >



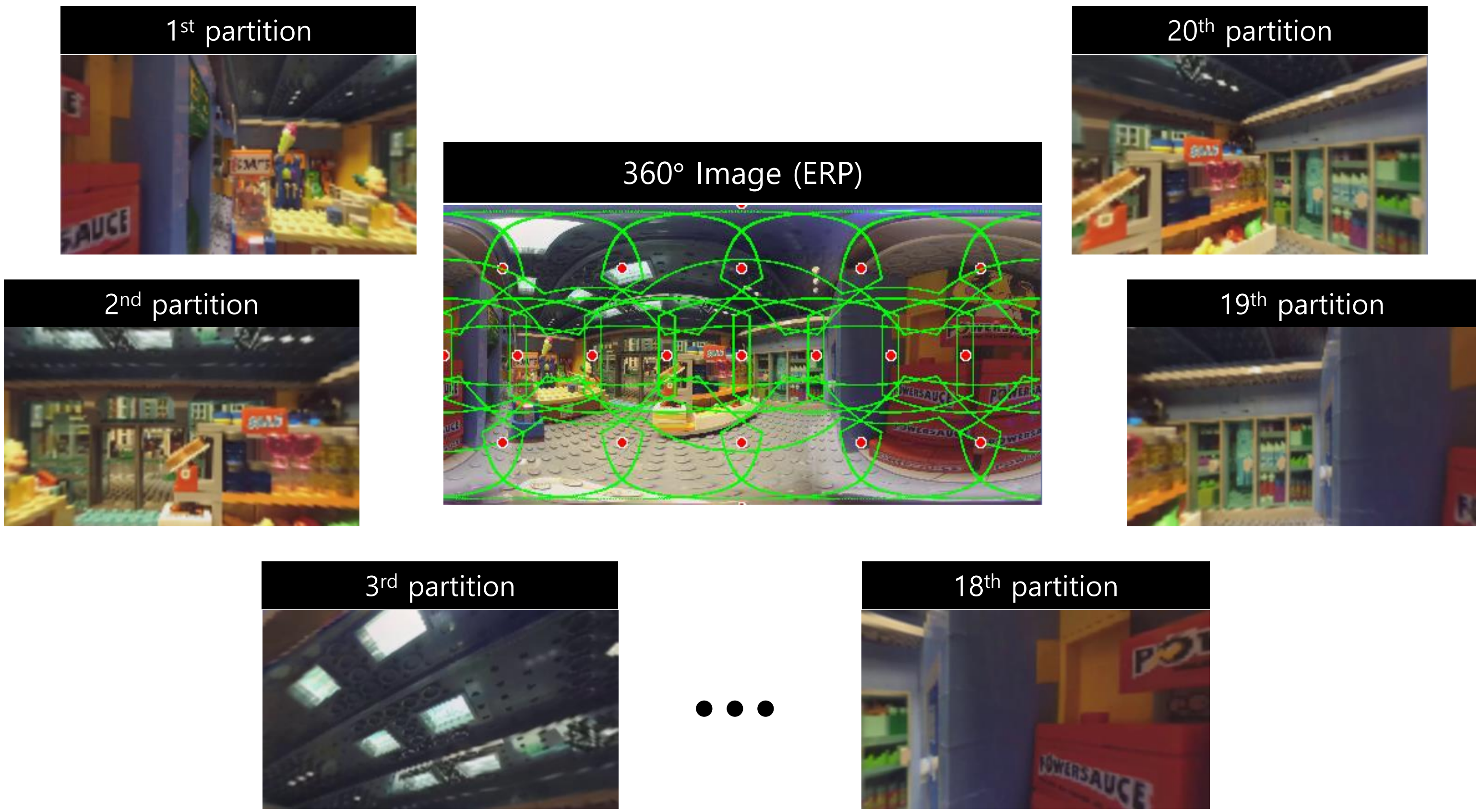
< Optical Flow >

[2] B. Alexe, T. Deselaers, and V. Ferrari. "Measuring the objectness of image windows." IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012.
[3] M. –M. Chang, T.-L. Liu, H.-T. Chen, and Y.-Y. Chuang. "BING: Binarized normed graidents for objectness estimation at 300fps." Conference on Computer Vision and Pattern Recognition, 2014.
[4] C.L. Zitnick and P. Dollar, "Edge boxes: Locating object proposals from edges." European Conference on Computer Vision, 2014.
[5] X. Cui, Q. Liu, and D. Metaxas. "Temporal spectral residual: fast motion saliency detection." International Conference on Multimedia, 2009.

Automatic Viewpoint Selection

Step1) Image Partitioning

- To reduce **undesirable distortion**, in 360° image (Equirectangular Projection, ERP), divide the image into several perspective images.



< Image Partition >

Automatic Viewpoint Selection

Step2) Scene Saliency Analysis (1/4)

- Scene Saliency: Salient region based on object's **EXISTENCE** and **MOTION**
- **Appearance saliency (Object existence probability)**

$$\boldsymbol{S}_{appear} = \frac{1}{\boldsymbol{Z}_{appear}} \sum_{i}^{N} score_{OP}(i)$$

$\boldsymbol{S}_{appear}$: Appearance Saliency

$score_{OP}(i)$: i^{th} object proposal's objectness score

$\boldsymbol{Z}_{appear}$: Normalization Factor

N: Number of object proposals

< Appearance Saliency Estimation Procedure >

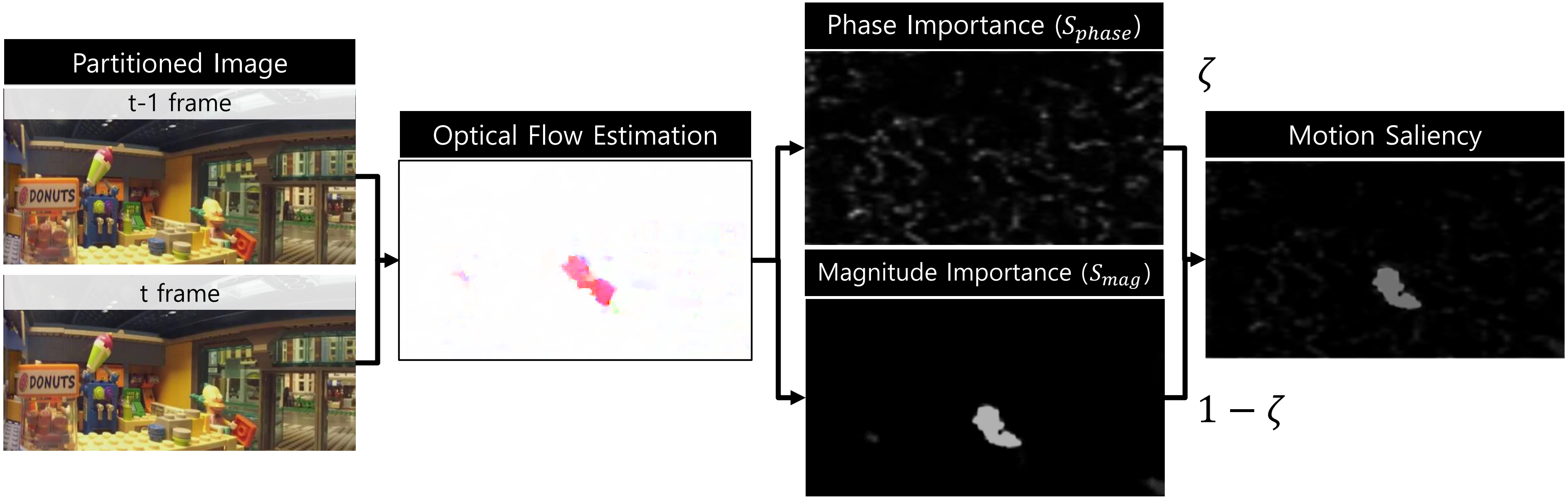
Automatic Viewpoint Selection

Step2) Scene Saliency Analysis (2/4)

- Scene Saliency: Salient region based on **OBJECT EXISTENCE** and **MOTION**
- **Motion Saliency (Movement Existence)**

$$S_{motion} = \zeta \cdot S_{phase} + (1 - \zeta) \cdot S_{mag}$$

S_{motion}: Motion saliency S_{phase}: Phase saliency S_{mag}: Magnitude saliency



< Motion Saliency Estimation Procedure >

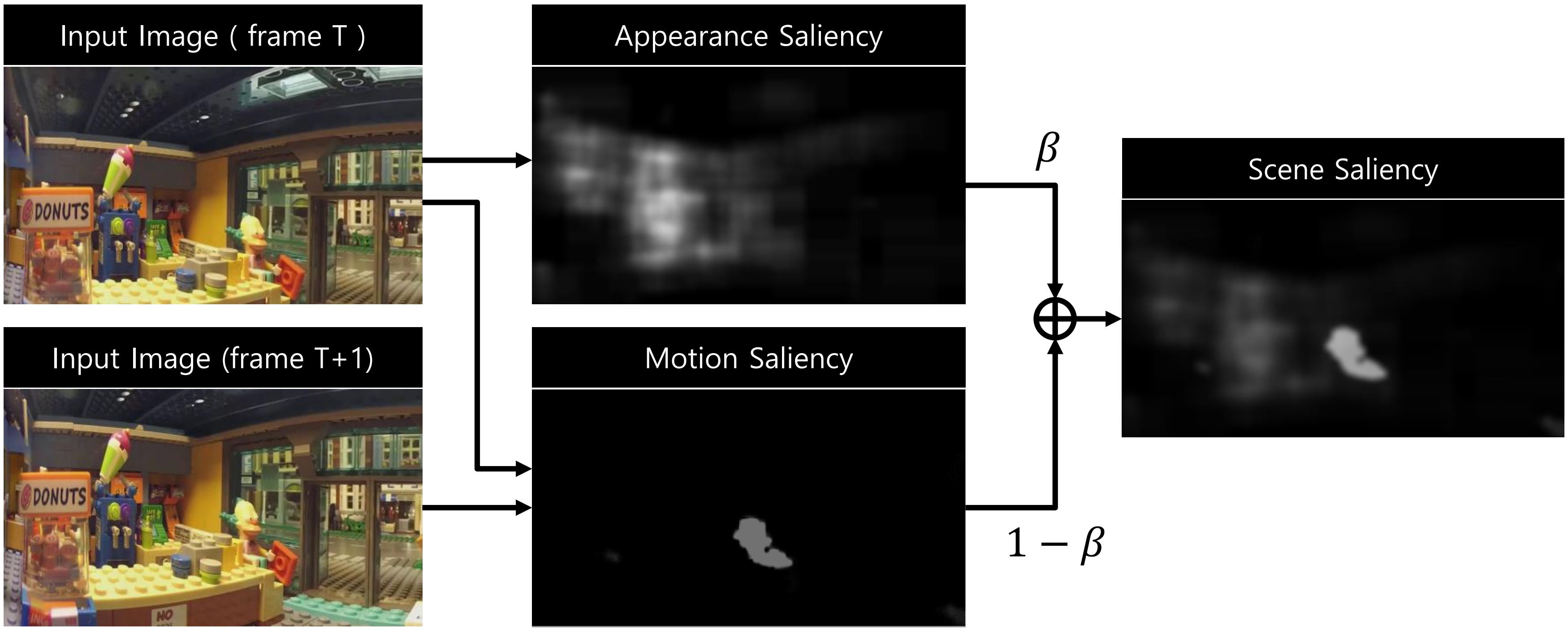
Automatic Viewpoint Selection

Step2) Scene Saliency Analysis (3/4)

- Scene Saliency: Salient region based on **OBJECT EXISTENCE** and **MOTION**

$$\boldsymbol{S}_{scene} = \beta * \boldsymbol{S}_{appear} + (1 - \beta) * \boldsymbol{S}_{motion}$$

$\boldsymbol{S}_{scene}$: Scene saliency $\boldsymbol{S}_{appear}$: Appearance saliency $\boldsymbol{S}_{motion}$: Motion saliency

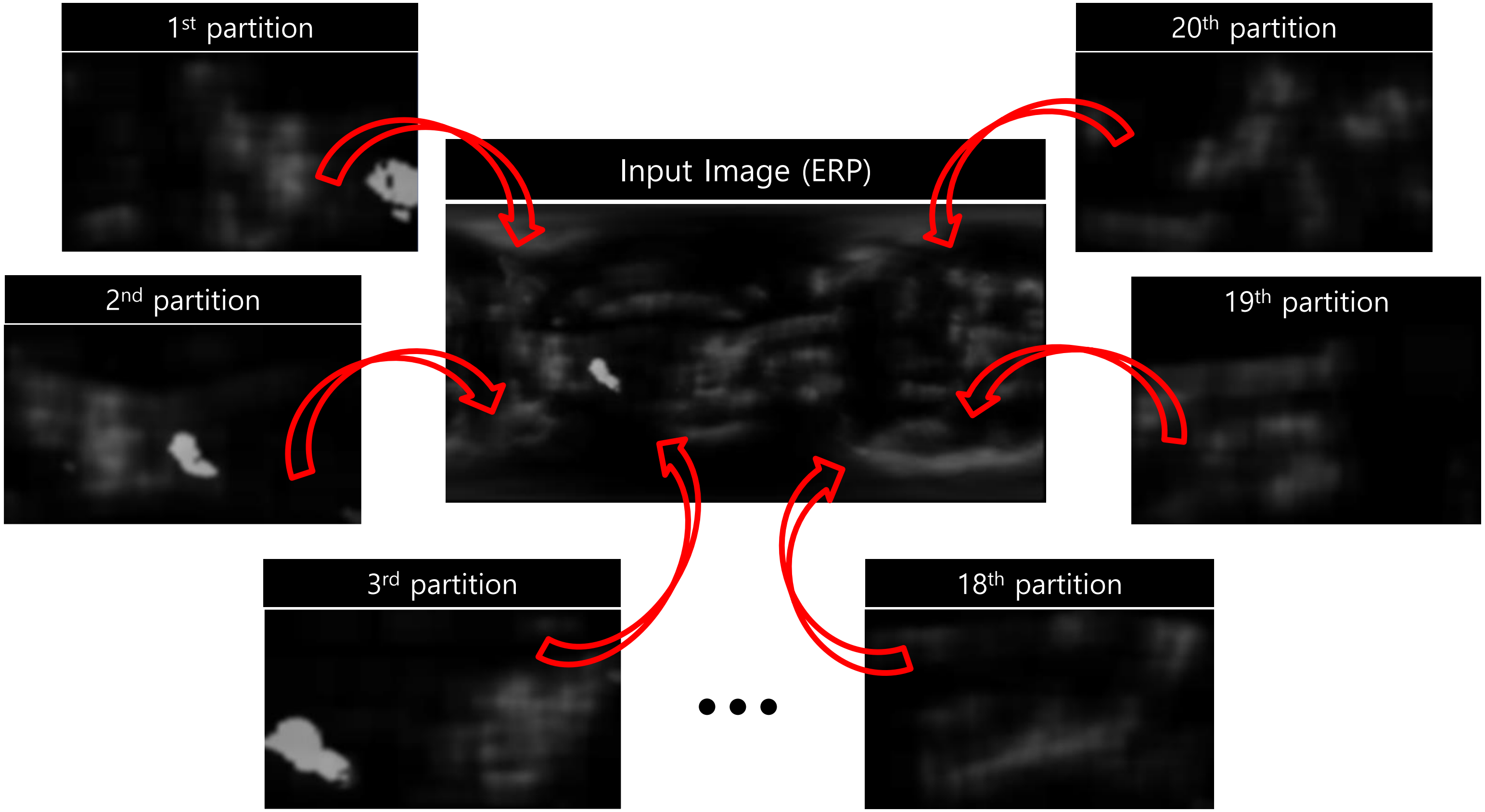


< Scene Saliency Estimation Procedure >

Automatic Viewpoint Selection

Step2) Scene Saliency Analysis (4/4)

- Do backward mapping the partitioned images into original one.



< Backward Mapping >

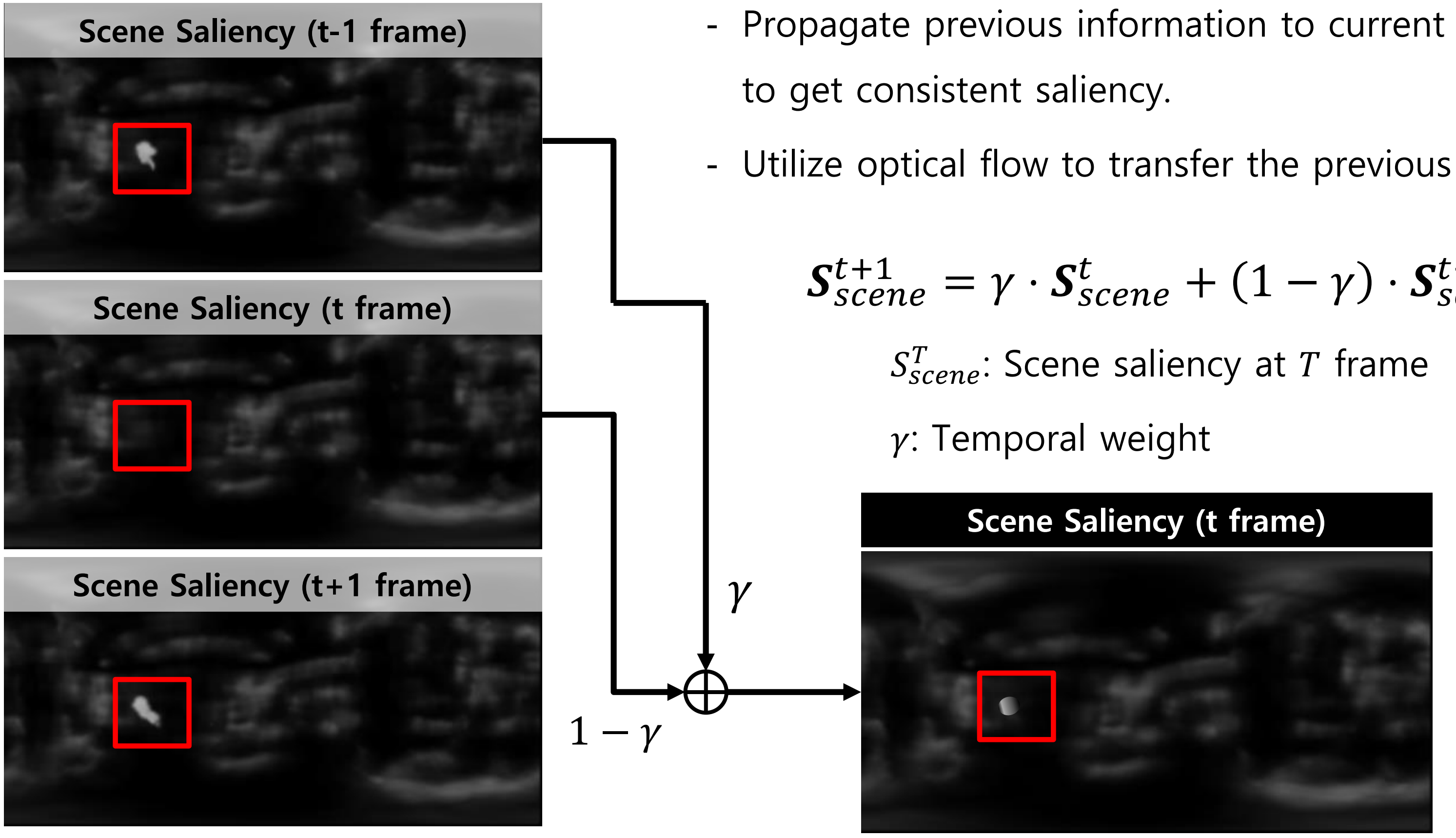
Automatic Viewpoint Selection

- **Scene Saliency Estimation Result**

Automatic Viewpoint Selection

Step3) Temporally Consistent Scene Saliency Estimation

- **Abruptly change** of luminance or motion of objects causes **inconsistent saliency**.



Example of inconsistent scene saliency

Example of temporally consistent scene saliency

- Propagate previous information to current frame to get consistent saliency.
- Utilize optical flow to transfer the previous saliency.

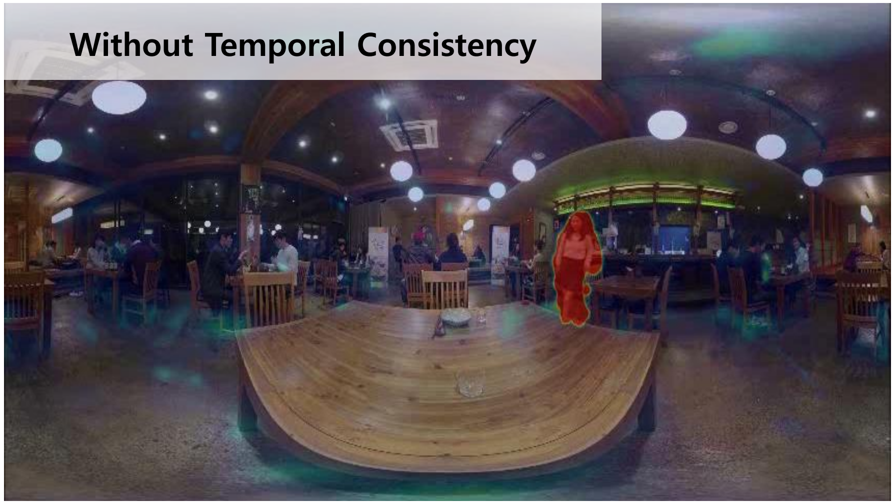
$$\boldsymbol{S}_{scene}^{t+1} = \gamma \cdot \boldsymbol{S}_{scene}^{t} + (1-\gamma) \cdot \boldsymbol{S}_{scene}^{t+1}$$

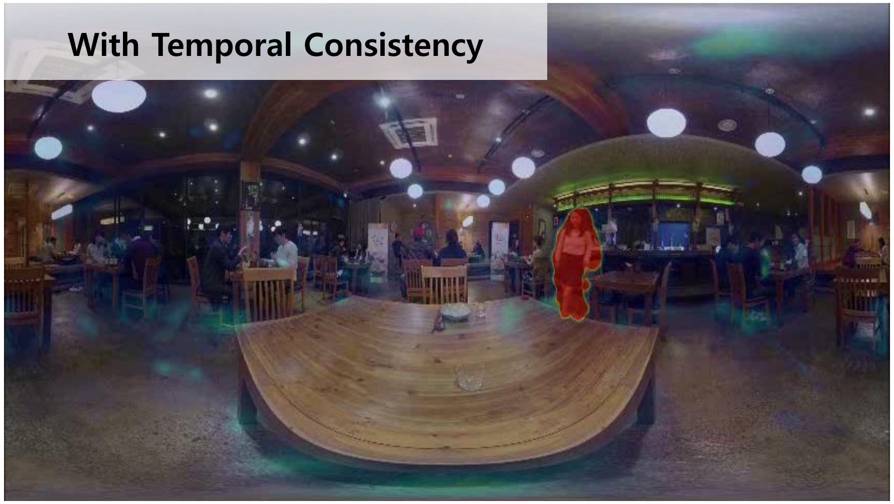
S_{scene}^{T}: Scene saliency at T frame

γ: Temporal weight

Automatic Viewpoint Selection

Temporally Consistent Scene Saliency Estimation





Automatic Viewpoint Selection

Step4) Viewpoint Selection

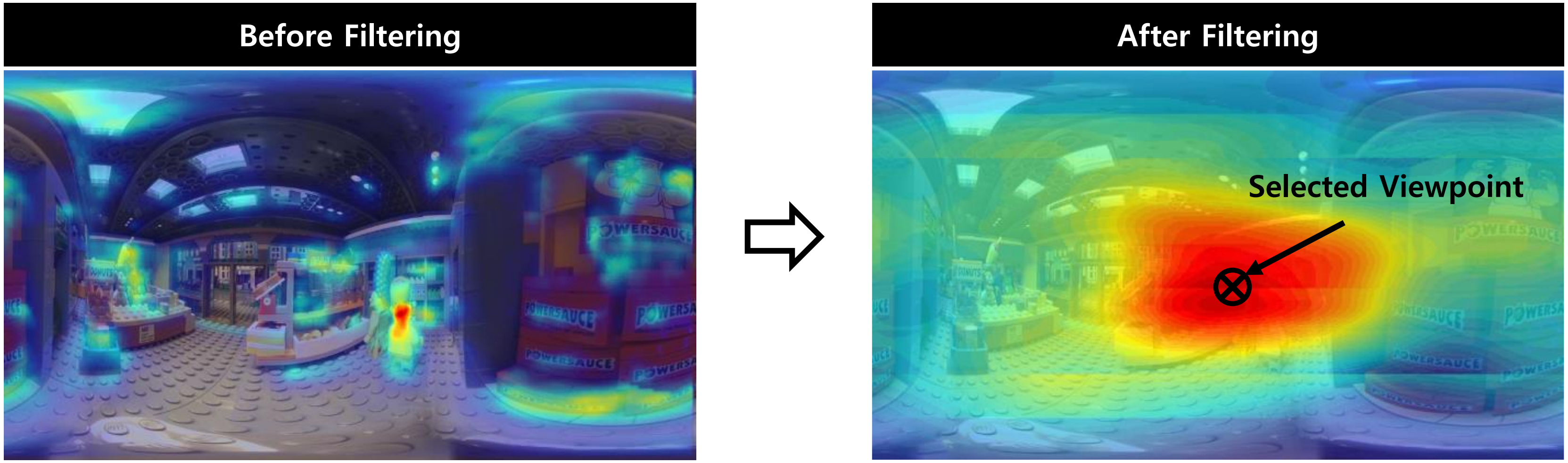
- Goal: Select the viewpoint that contains salient parts as many as possible.
  - To do this, filter estimated saliency map, then choose the peak as the viewpoint.

$$\boldsymbol{S}_{scene}^{filtered} = \mathcal{F}_{\text{avg}} * \boldsymbol{S}_{scene}$$

$\boldsymbol{S}_{scene}^{filtered}$: Filtered Scene Saliency Map

$\mathcal{F}_{\text{avg}}$: Average Filter

$\boldsymbol{S}_{scene}$: Scene Saliency



< Comparison on scene saliency with/without filtering >

Demo

Automatic Viewpoint Selection on 360 Videos