5D Stabilization through Sensor Vision Fusion

Binnan Zhuang[1], Dongwoon Bai[2], and Juangwon Lee[2]

1. Xsense.ai
2. Samsung Semiconductor, Inc.

5D Stabilization through Sensor Vision Fusion

Outlines:

5D Stabilization Overview
From 3D stabilization to 5D stabilization
Efficient Path Optimization
Performance Comparison
Conclusions and Future Directions

Video Stabilization Techniques

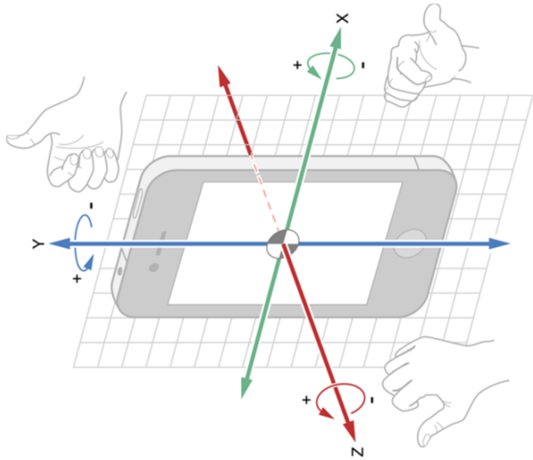
- **Optical Image Stabilization (OIS):**
  - **Measure instantaneous camera movements through inertial sensors.**
  - **Compensate camera oscillation before image is projected.**
  - **Compensation is achieved through mechanically moving the lens or sensor.**
  - **Capable of filtering out high frequency motion jitter with small magnitude.**

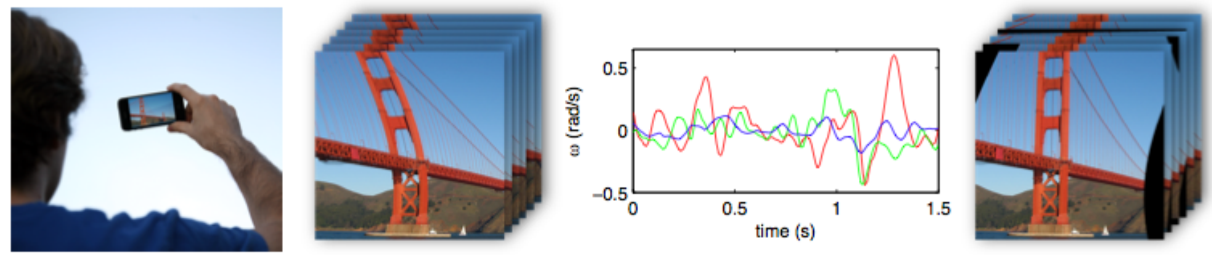
- **Digital Image Stabilization (DIS):**
  - **Estimate a camera motion trajectory.**
  - **Decide the smooth motion trajectory through camera path smoothing.**
  - **Compensation is achieved through digital image warping.**
  - **Adapt to dynamic camera motion and achieve better smoothing using trend filtering.**
  - **The proposed 5D stabilization is a DIS approach.**

Existing DIS Methods

▪ Gyro based 3D stabilization

- [Karpenko, et al, `2011]
- Widely used for real time video stabilization on smart phones.



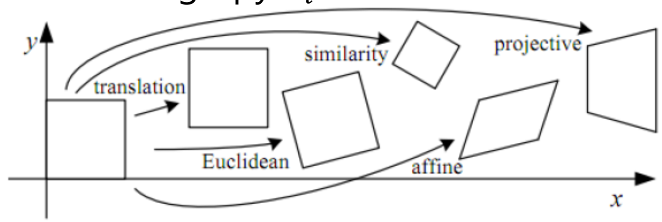


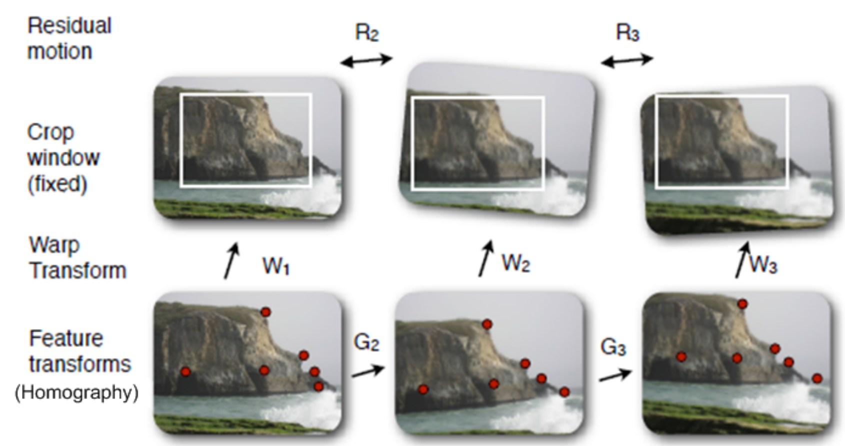
(a) (b) (c) (d)

▪ Vision based stabilization using homography

- [Grundmann, et al, `2011]
- Too complex for real time application.
- Performance depends on feature tracking quality.

Homograpy G_i:



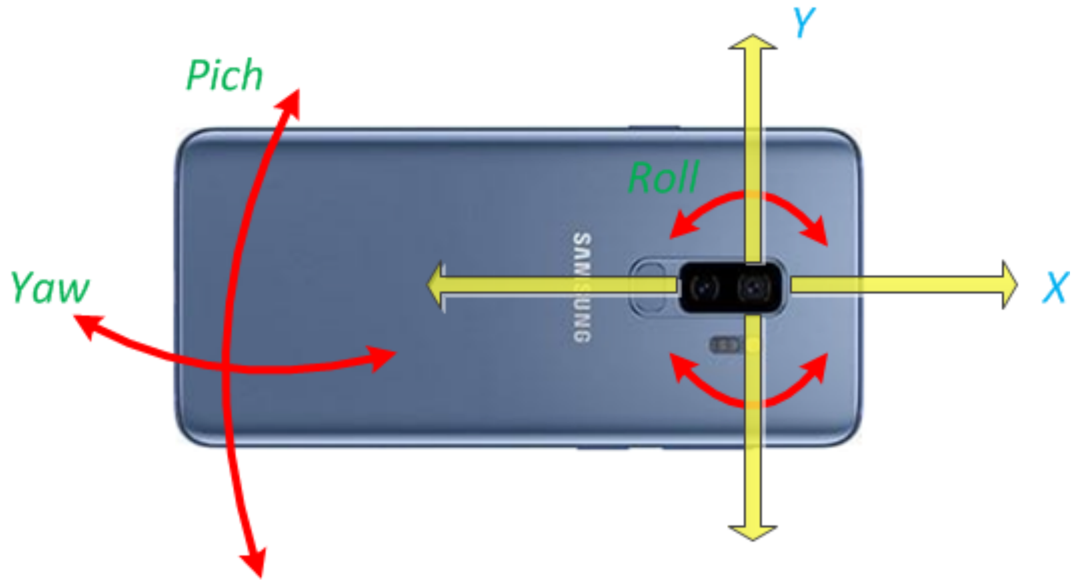


Motivation for Sensor Vison Fusion

- **Problems with pure gyro based or vision based solution**
  - **Gyro based methods can only compensate 3D rotations, which will suffer in scenes with highly dynamic translation.**
- **Intuition of 5D stabilization:**
  - **Obtain precise 3D rotation estimates using a gyroscope.**
  - **Estimate the effect of 3D translation from MVs, without depth information.**

5D Stabilization Overview

- **Sensor vision fusion:**
  - Sensor: gyroscope.
  - Vision: motion vectors (MVs) obtained from consecutive frames.
- **5D video stabilization: 3D rotation + residual 2D translation**
  - 3D rotation is measured from a gyroscope.
  - Residual 2D translation is estimated from MVs.

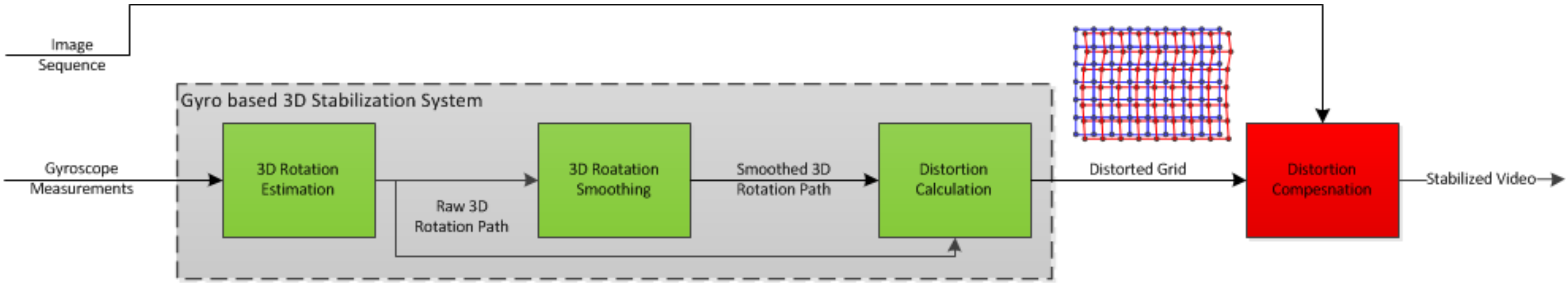


5D Stabilization through Sensor Vision Fusion

Outlines:

5D Stabilization Overview
From 3D stabilization to 5D stabilization
Efficient Path Optimization
Performance Comparison
Conclusions and Future Directions

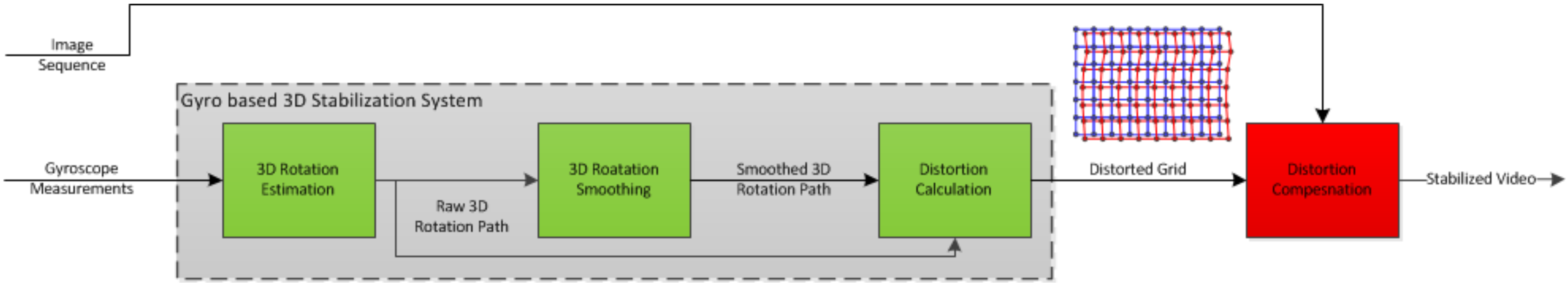
Gyro based 3D Stabilization

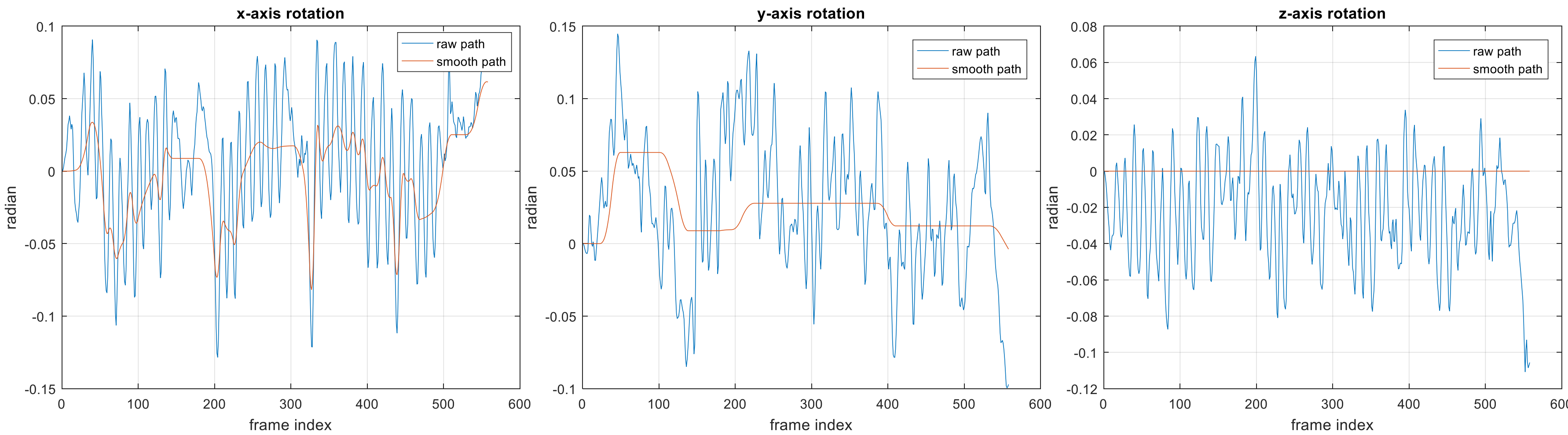


- **The raw 3D rotation path consists of 3×1 rotation vectors, representing the accumulated camera rotations from the initial frame.**

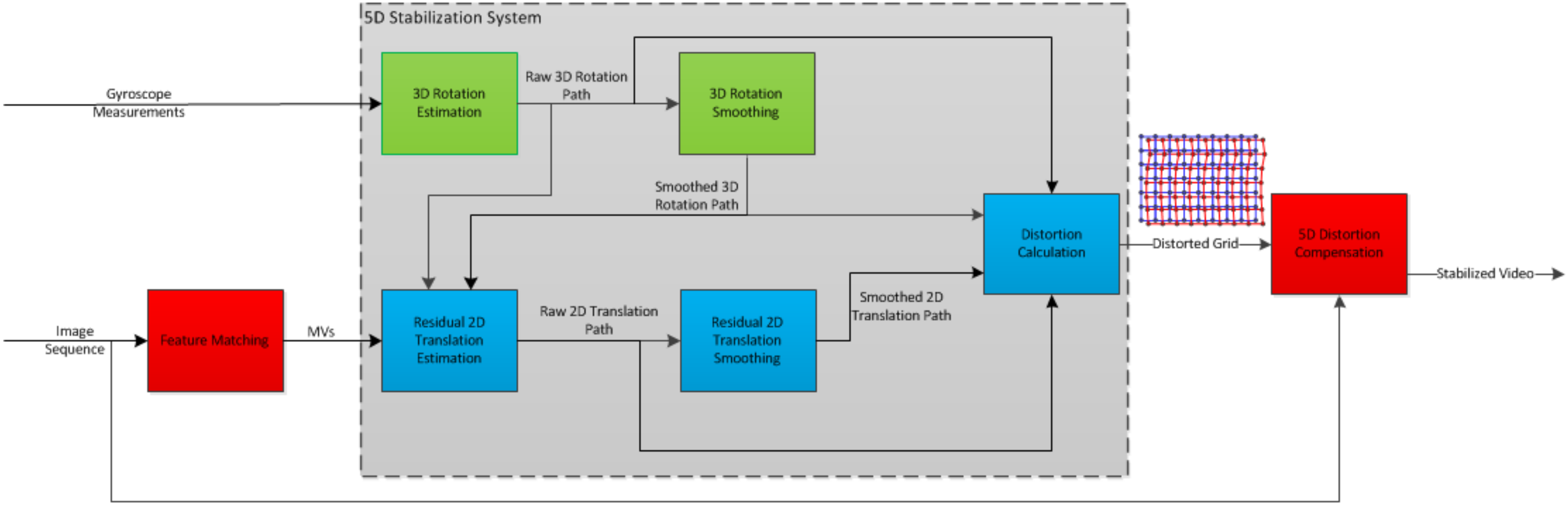
- **The smoothed 3D rotation path can be obtained by solving corresponding path optimization problems, which indicates the stabilized camera rotations.**

Gyro based 3D Stabilization





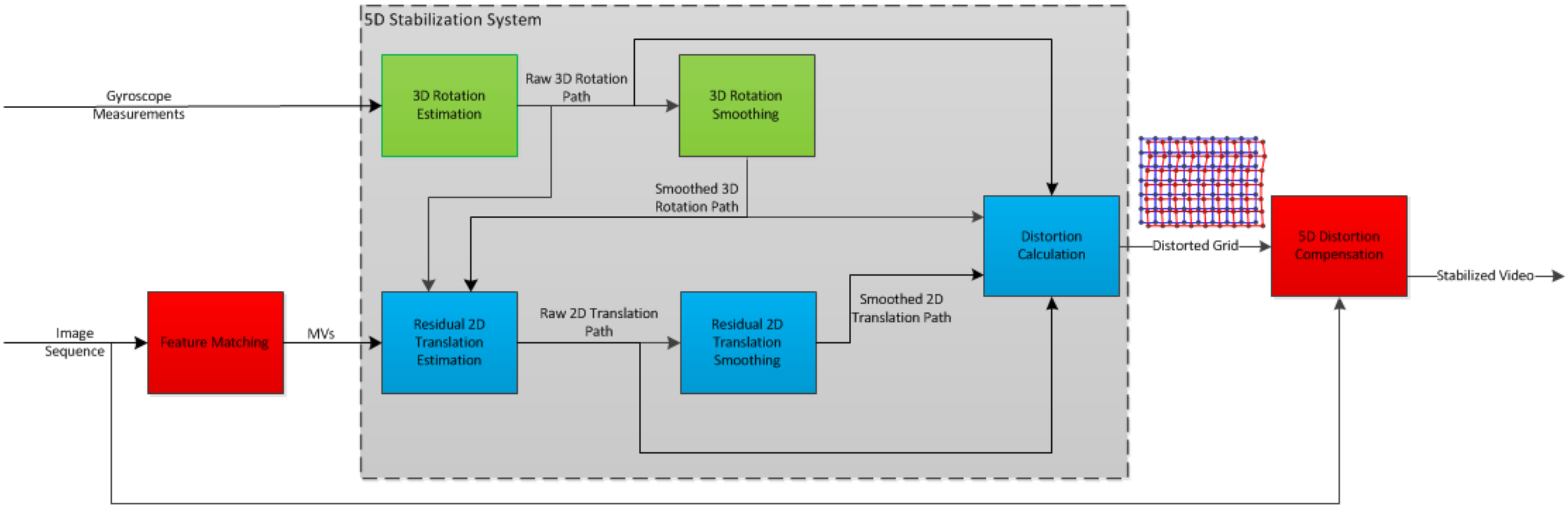
5D Stabilization System

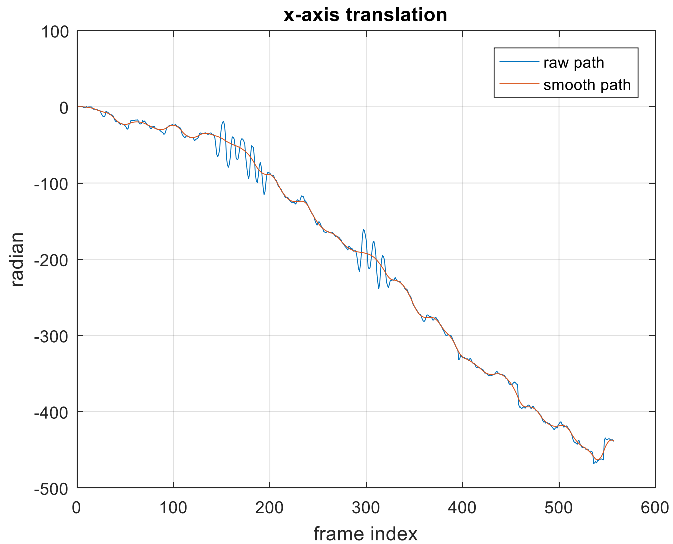


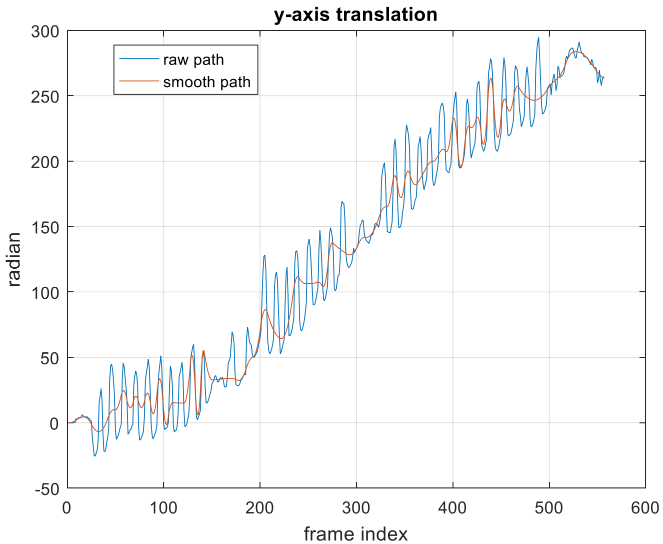
- **The raw 2D translation path consists of 2×1 translation vectors, representing the accumulated translation (within image plane) from the initial frame.**

- **The smoothed 2D translation path is again obtained by solving corresponding path optimization problem.**

5D Stabilization System

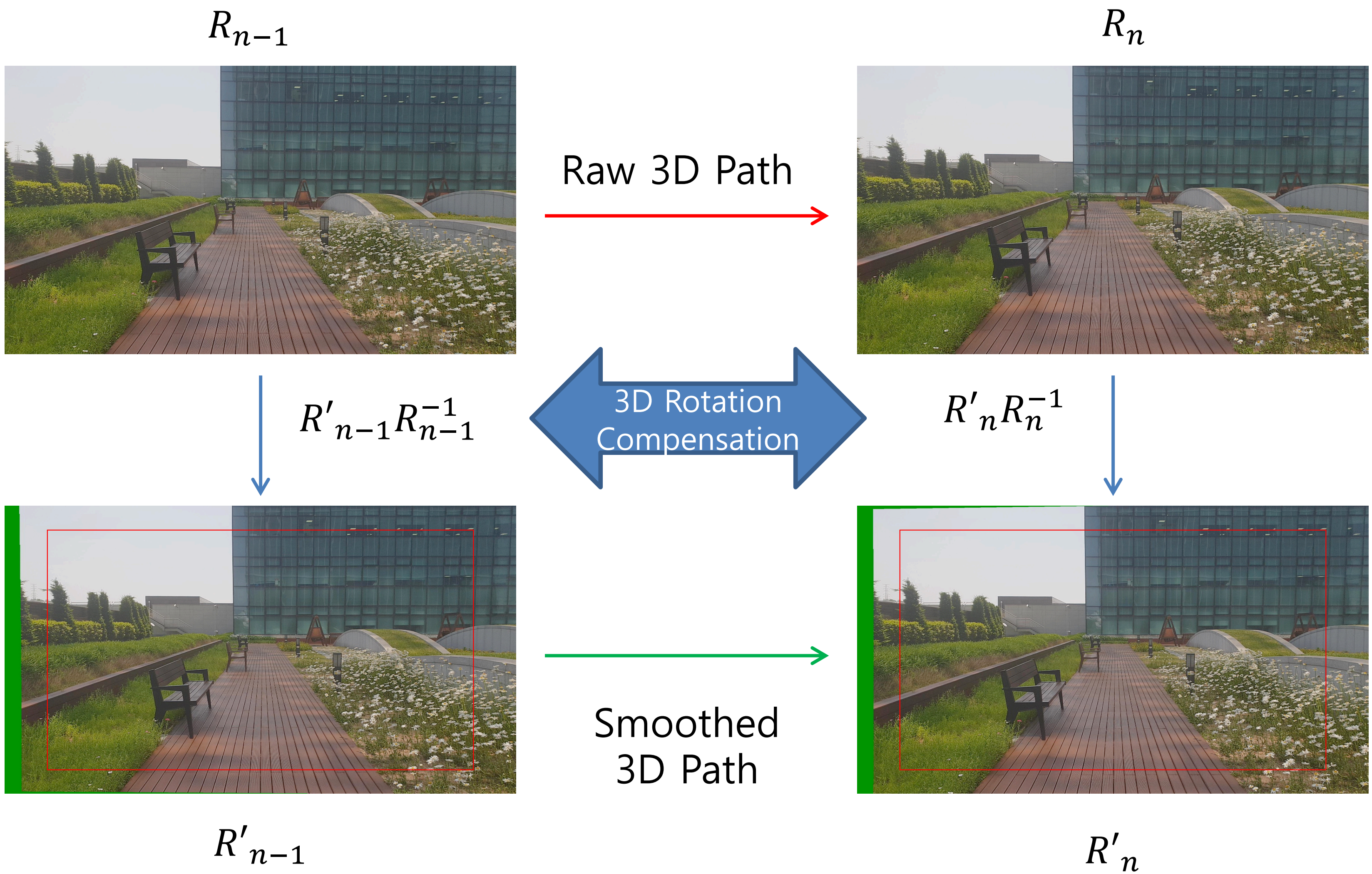






3D Compensation

- **Denote the raw and stabilized 3D rotations from the initial frame to frame n as R_n and R'_n, which are 3×3 rotation matrices.**



Residual 2D Translation Estimation

- **We want to characterize the remaining impact due to 3D translation after 3D rotation compensation.**

- **The solution is residual 2D translation estimation using MVs.**

- **Denote the mth MV from frame $n-1$ to frame n as:**

$$(x_{n-1}^m, y_{n-1}^m) \rightarrow (x_n^m, y_n^m)$$

- **The residual translation calculated with pose alignment:**

$$\begin{bmatrix} \tilde{x}_{n-1}^m \\ \tilde{y}_{n-1}^m \\ \tilde{z}_{n-1}^m \end{bmatrix} = K R_n' R_{n-1}^{-1} K^{-1} \begin{bmatrix} x_{n-1}^m \\ y_{n-1}^m \\ 1 \end{bmatrix}, \quad \begin{bmatrix} \tilde{x}_n^m \\ \tilde{y}_n^m \\ \tilde{z}_n^m \end{bmatrix} = K R_n' R_n^{-1} K^{-1} \begin{bmatrix} x_n^m \\ y_n^m \\ 1 \end{bmatrix}$$

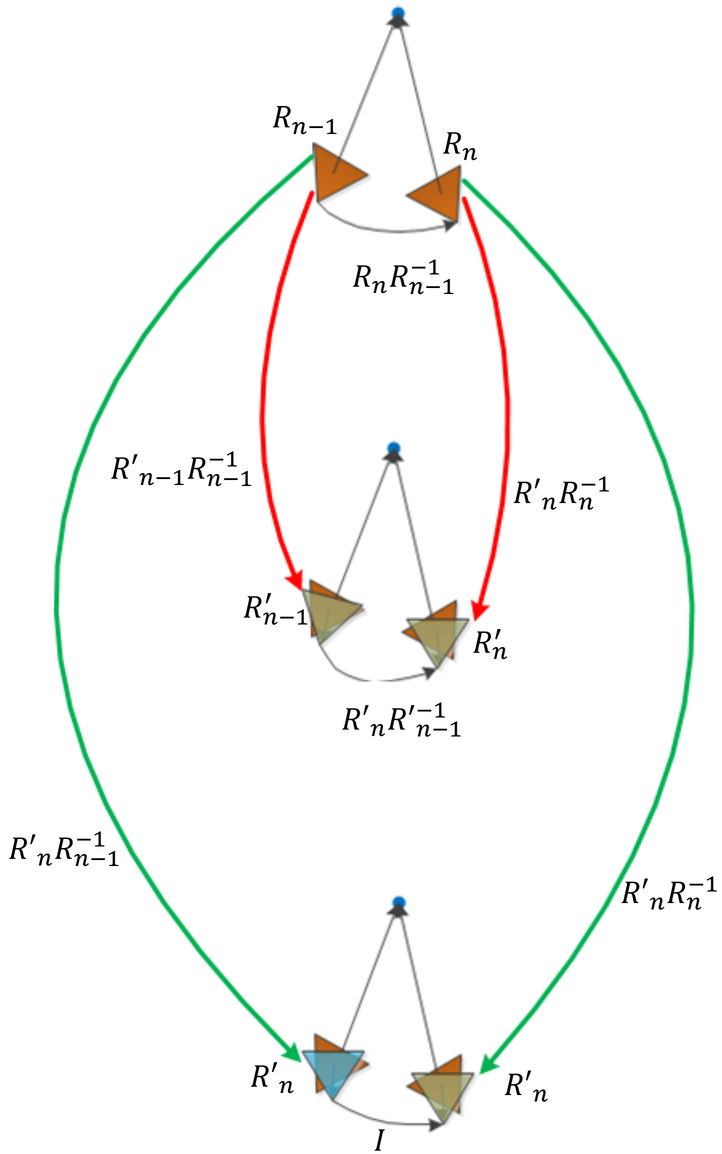
$$\Delta T_{n-1\rightarrow n}^m = \begin{bmatrix} \tilde{x}_n^m / \tilde{z}_n^m - \tilde{x}_{n-1}^m / \tilde{z}_{n-1}^m \\ \tilde{y}_n^m / \tilde{z}_n^m - \tilde{y}_{n-1}^m / \tilde{z}_{n-1}^m \end{bmatrix},$$

where K is the camera intrinsic matrix.

- **The inter-frame residual 2D translation is:**

$$\Delta T_{n-1\rightarrow n} = \frac{1}{M} \sum_{m=1}^{M} \Delta T_{n-1\rightarrow n}^m$$

Intuition behind Pose Alignment



Raw camera pose

After 3D rotation compensation

- The residual 2D translation estimated after 3D rotation compensation (without pose alignment) will incorrectly treat $R'_n R'^{-1}_{n-1}$ as part of the residual 2D translation.
- The 2D translation compensation based on such estimates will contaminate the stabilized 3D rotation path.
- Significant performance degradation during large turns.

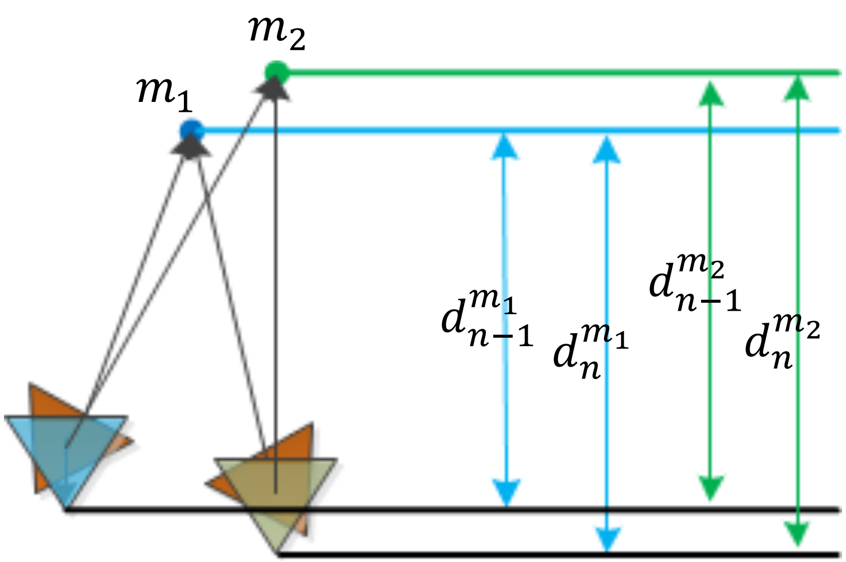
After pose alignment

- The residual 2D translation estimated with pose alignment will capture the end effect due to pure 3D translation within the image plane.
- The raw 2D path can be directly obtained as: $T_n = \sum_{i=1}^{n} \Delta T_{n-1 \to n}$.
- Alignment to the stabilized 3D rotation in the current frame is also important, because the corresponding residual 2D translation represents the actual translation jitter after 3D rotation compensation.

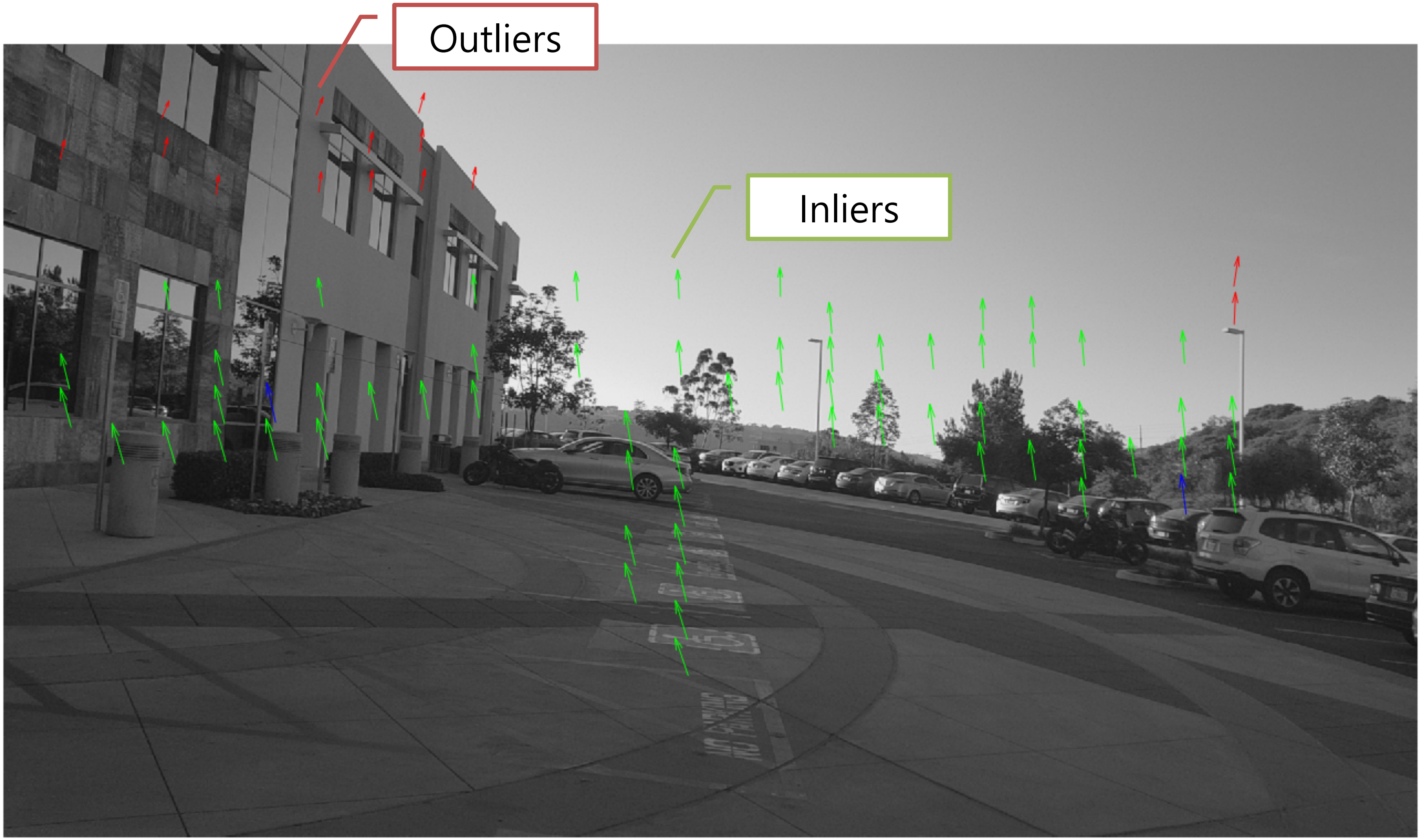
Depth Related Approximation

Samsung Confidential

- **The end effect of a pure 3D translation depends on depth.**
- **Averaging over MVs relies on two approximations:**
  - The translation along z-axis is much smaller than the depth of the object point, i.e., $|d_n^m - d_{n-1}^m| \ll d_{n-1}^m$.
  - The depths of different object points are also close, i.e., $d_{n-1}^{m_1} \approx d_{n-1}^{m_2}$.
  - Mesh-based residual 2D translation estimation similar to [Liu, et al, `2013] can be used to handle depth variation.



Residual 2D MVs



5D Stabilization through Sensor Vision Fusion

Samsung Confidential
SAMSUNG ELECTRONICS
System LSI Business

Supplementary Slides:

Camera Path Optimization
5D Stabilization with Motion Prediction

Path Optimization Problem

- **$L1$ path optimization [Grundmann, et al, 2011] :**

$$\min_{\boldsymbol{x}} \; w_0\|\boldsymbol{x}-\boldsymbol{y}\|_1^1 + \sum_{i=1}^{3}\|\boldsymbol{D}_i\boldsymbol{x}\|_1^1$$

$$\text{s.t.} \;\; \boldsymbol{l} \leq \boldsymbol{x}-\boldsymbol{y} \leq \boldsymbol{u}$$

$$x_{n-a_1+i} = x^*_{n-a_1+i}, \qquad i = 0,\cdots,a_1-1$$

  - $\boldsymbol{y} = \left[y_{n-a1},\cdots,y_n,\cdots,y_{n+a_2}\right]$ is the raw path, $\boldsymbol{x}$ is the smoothed path to be optimized.
  - The box constraint is to guarantee that the stabilized image will cover the entire cropping window, where $\boldsymbol{l}$ and $\boldsymbol{u}$ are dynamically calculated.
  - The equality constraint ensures the pervious optimized values are not changed.
  - In frame n, only x_n^* corresponding to the current frame is used in the stabilized path.

Efficient QP Solution

- **$L2$ path optimization:**

$$\min_{\boldsymbol{x}} \; w_0 \|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \sum_{i=1}^{3} w_i \|\boldsymbol{D}_i \boldsymbol{x}\|_1^1$$
$$\text{s.t.} \;\; \boldsymbol{l} \le \boldsymbol{x}-\boldsymbol{y} \le \boldsymbol{u}$$

  - $\boldsymbol{y} = \left[y_{n-a1}, \cdots, y_n, \cdots, y_{n+a_2}\right]$ is the raw path, $\boldsymbol{x}$ is the smoothed path to be optimized.
  - The box constraint guarantees that the stabilized image covers the cropping window.
  - The equality constraint ensures the pervious optimized values are not changed.

Efficient Solution based on ADMM

- The path optimization problem can be converted to a Quadratic Programming (QP) problem through dual transform.

- The QP problem is solved by an iterative algorithm based on Alternating Direction Method of Multipliers (ADMM).
  - Utilizing the special structure of the problem, the ADMM update can be computed efficiently in closed form.

- The ADMM based QP solution achieves 73.5% and 52.1% run time reduction compared to solving the L2 optimization using a standard QP solver and solving the L1 optimization in [Grundmann, et al, `2011].

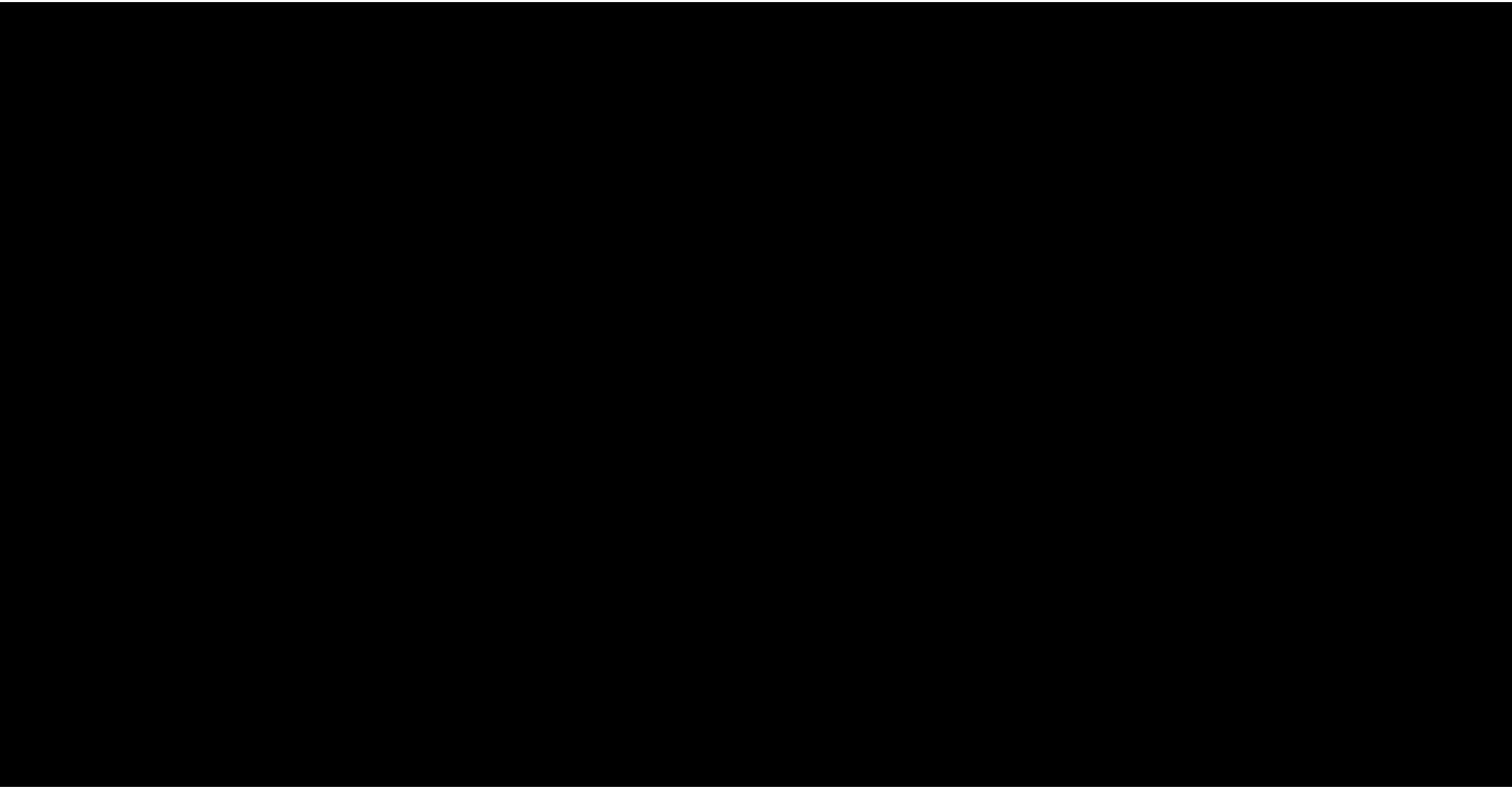
- The efficient path optimization solver allows us to prototype the 5D stabilization on a Galaxy S8 for 30 fps real-time video recording.

5D Stabilization through Sensor Vision Fusion

Outlines:

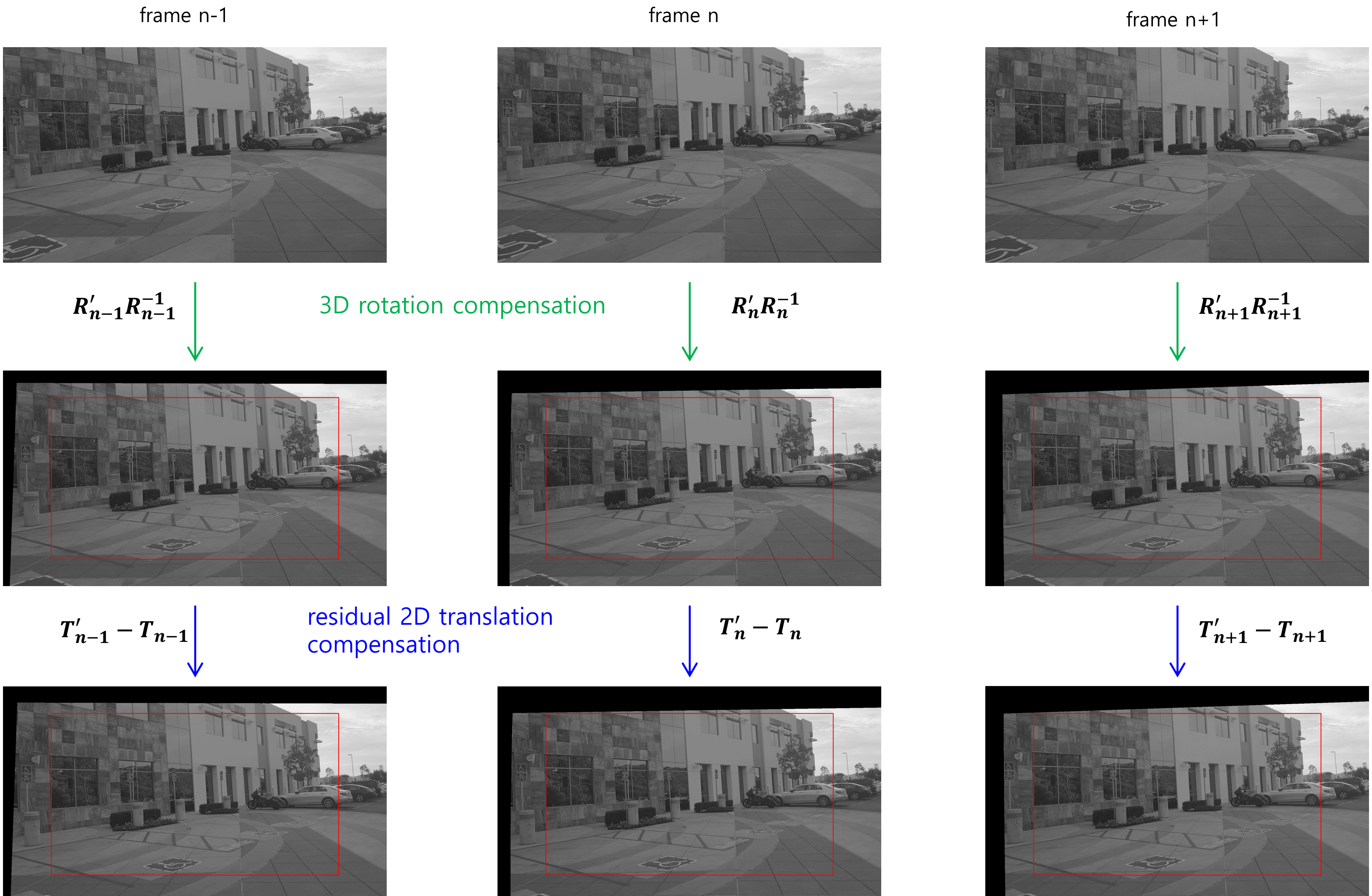
5D Stabilization Overview
From 3D stabilization to 5D stabilization
Performance Comparison
Conclusions and Future Directions

5D Stabilization vs. State-of-the-Art Solutions



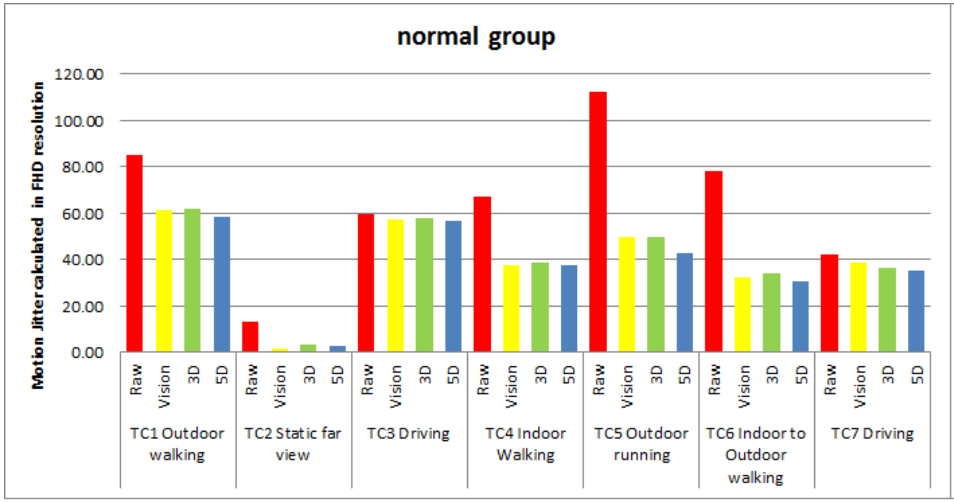
5D Stabilization vs. 3D Stabilization

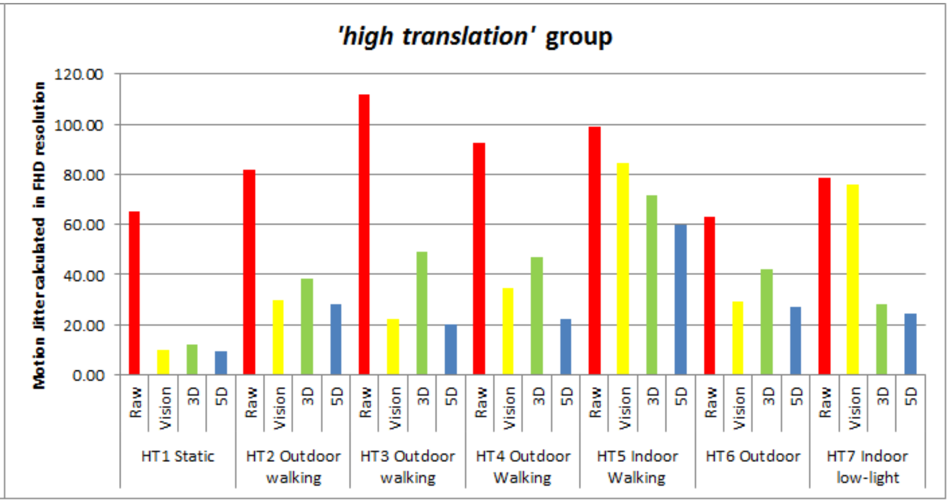
5D Compensation



Motion Jitter Comparison

- **Against state-of-the-art-solutions**





5D Stabilization through Sensor Vision Fusion

Outlines:

5D Stabilization Overview
System Architecture
2D Translation Estimation
Performance Comparison
Conclusions and Future Directions

Conclusions and Future Directions

- **Conclusions:**
  - **5D stabilization inherits the merits of both gyro and vision based video stabilization through sensor-vision fusion.**
  - **5D stabilization significantly improves the performance over 3D stabilization in scenes with high translation movements.**

- **5D stabilization for object of interest (OOI):**
  - **3D background stabilization + residual 2D OOI stabilization.**
  - **5D OOI stabilization for front facing camera video recording.**

- **Future directions:**
  - **6D stabilization (using depth sensor).**

Thank You