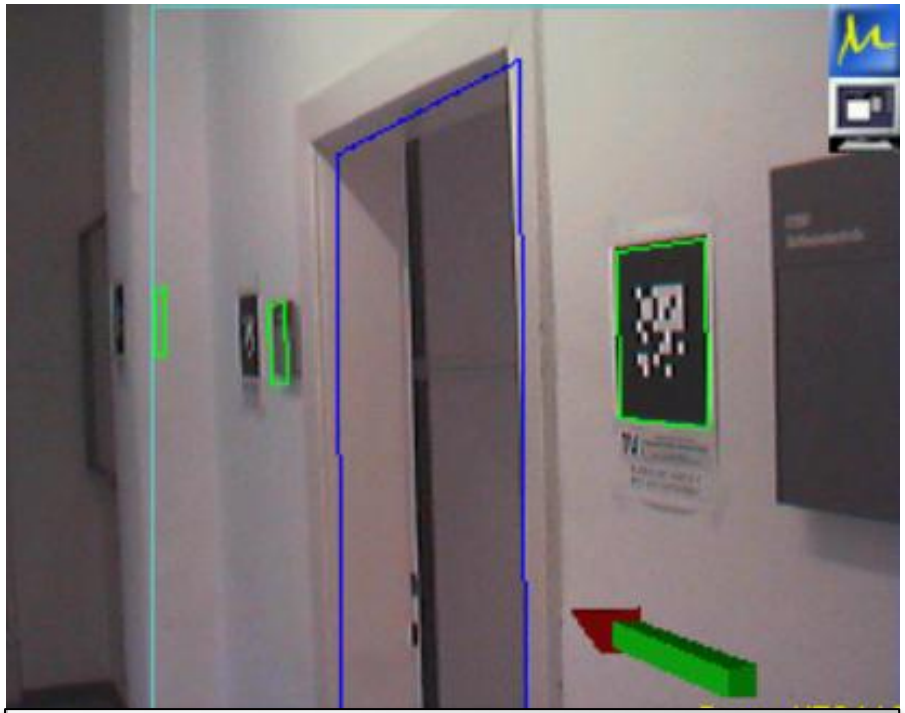
Photo-shoot localization for mobile AR based on registered images in virtualized reality models

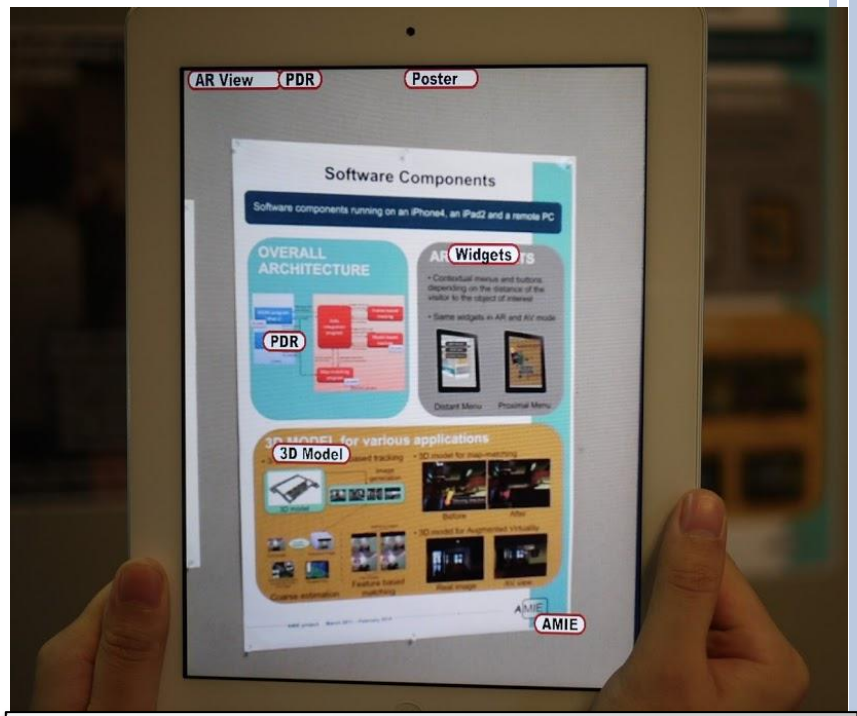
Jun Nishida [12], Koji Makita [2], Tomoya Ishikawa [2],
Takashi Okuma [2], Masakatsu Kourogi [2], Thomas Vincent [3],
Laurence Nigay [3], Jun Yamashita [1],
Hideaki Kuzuoka [1], Takeshi Kurata [12]

1. University of Tsukuba, Japan
2. Center for Service Research, AIST, Japan
3. Université Joseph Fourier, France

AR applications



D. Wagner, et al. “First steps towards handheld augmented reality” Proc, ISWC 2003.



Univ of Tsukuba, Demo in IDW/AD'12.

Camera position and orientation are needed for creating AR images

Research background ～localization methods～

Outdoor environment

- GPS（Global Positioning System）
  - Mainly used for navigation

Indoor environment

- Method using physical Infrastructure（WIFI, RFID）
  - Task : installation costs
- **Method using image processing**
  - Task : cost of creating the reference DB

In the indoor environment, estimating the position and orientation is dependent on the condition.

Relative and absolute localization

Relative localization

- Get only the movement from the reference position and orientation
- Computational cost
- Can update the position and orientation in real time information

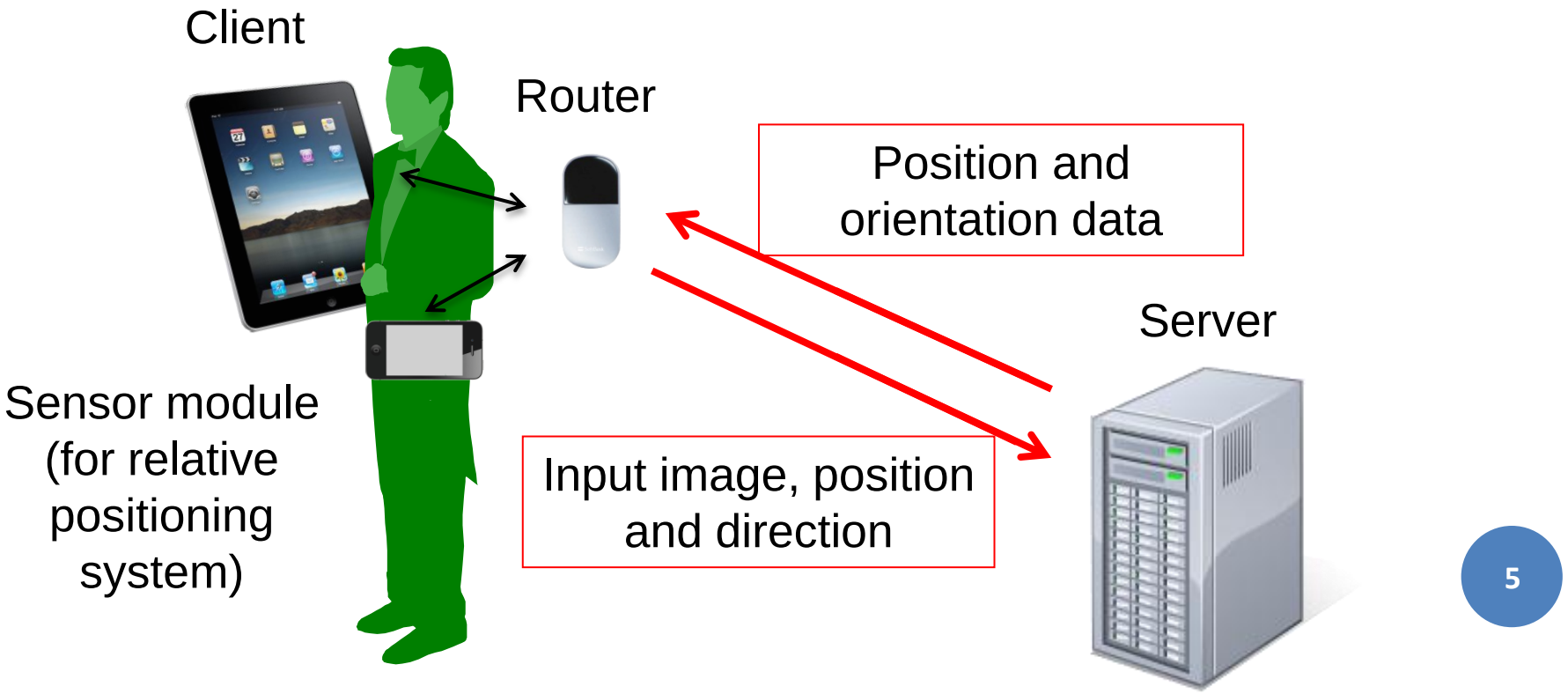
Absolute localization

- Get the absolute position and orientation
- Can be used to initialize or correction of relative positioning
- Difficult to update the position and orientation in real time

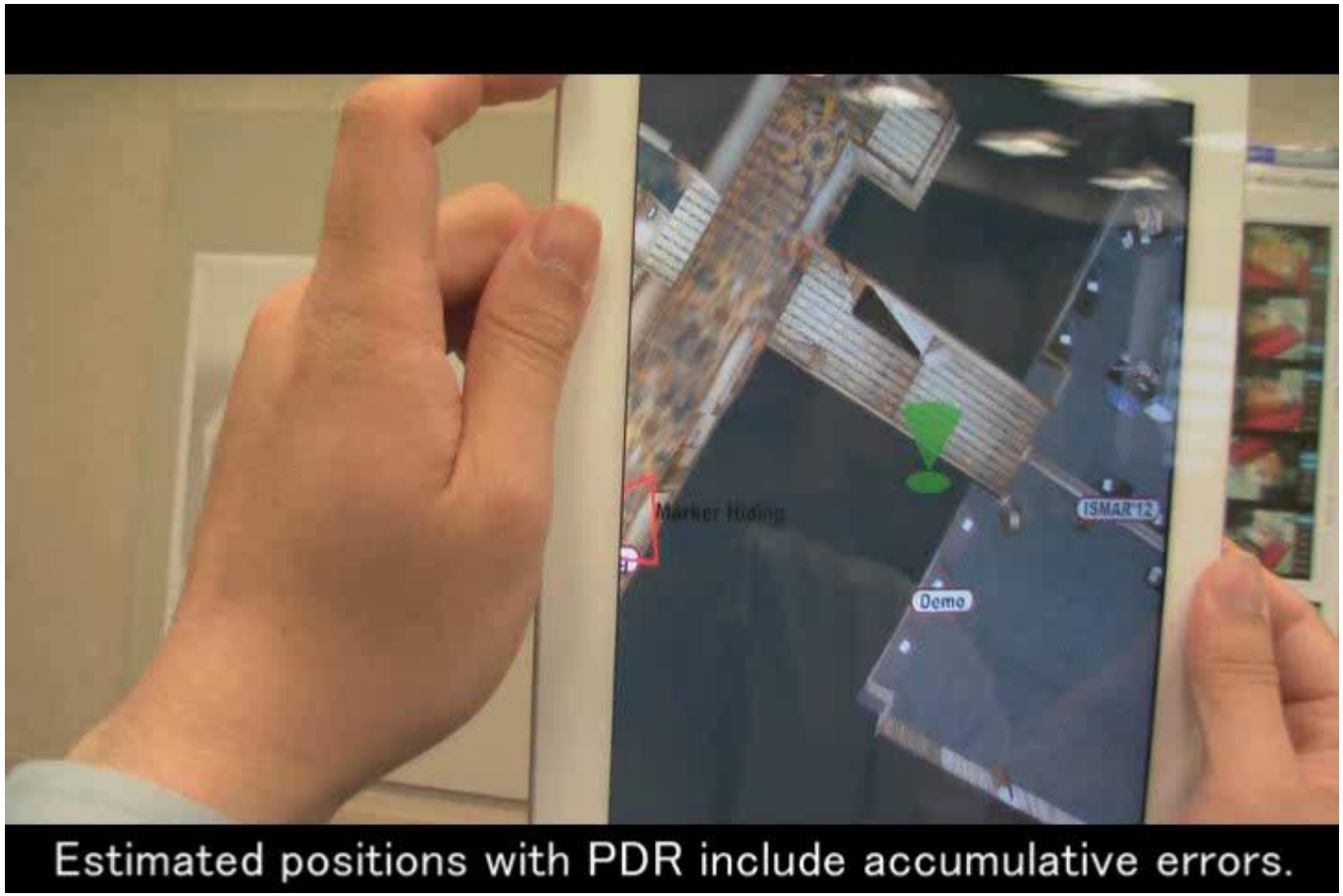
We construct the absolute positioning system based on registered images in virtualized reality model

System architecture for combining the relative positioning system

- Server-client system
- Always estimate the relative position and orientation using the relative positioning system (PDR)
- Absolute position and orientation are estimated in the server when the camera image is sent to the server



Appearance of the demonstration (in ISMAR2012)



Start without the position information

↓

Estimation of position and orientation using image processing

↓

Correction of the position and orientation

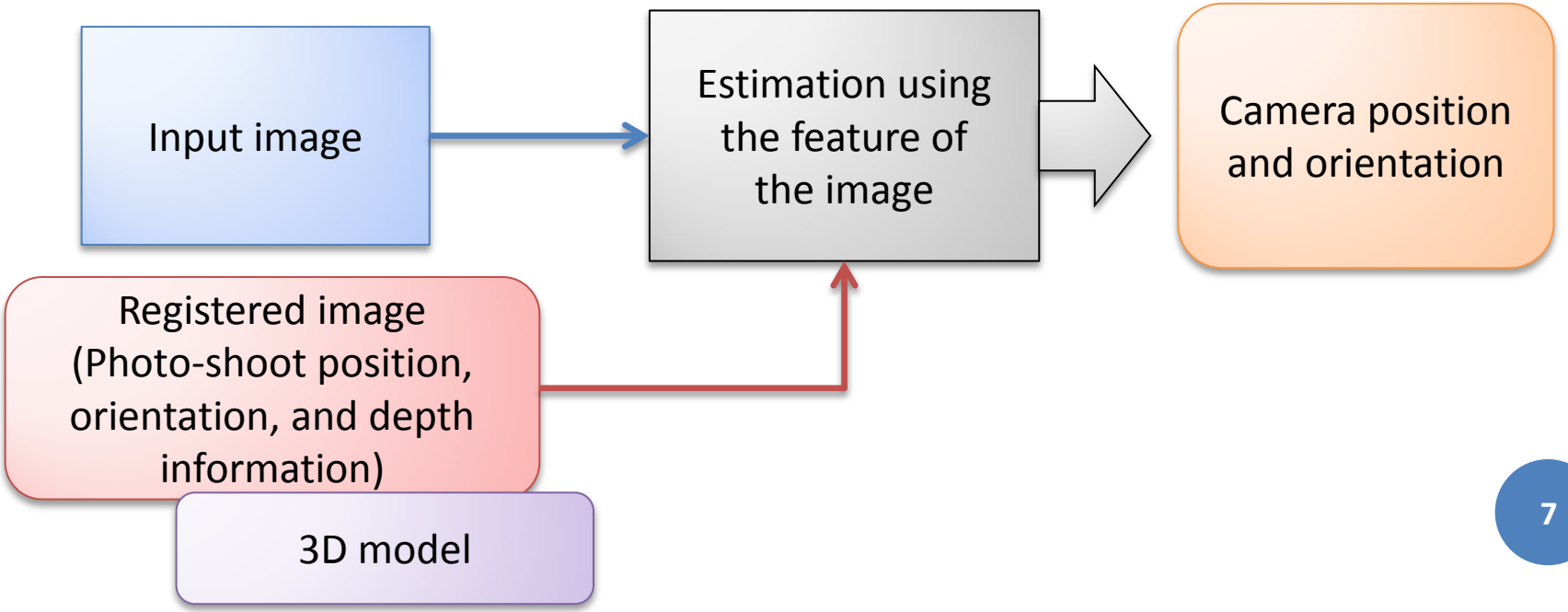
↓

The relative positioning using PDR system

Correspondence of image and 3D model

We compare the input image and the registered images using feature point, edge ,and surface information.

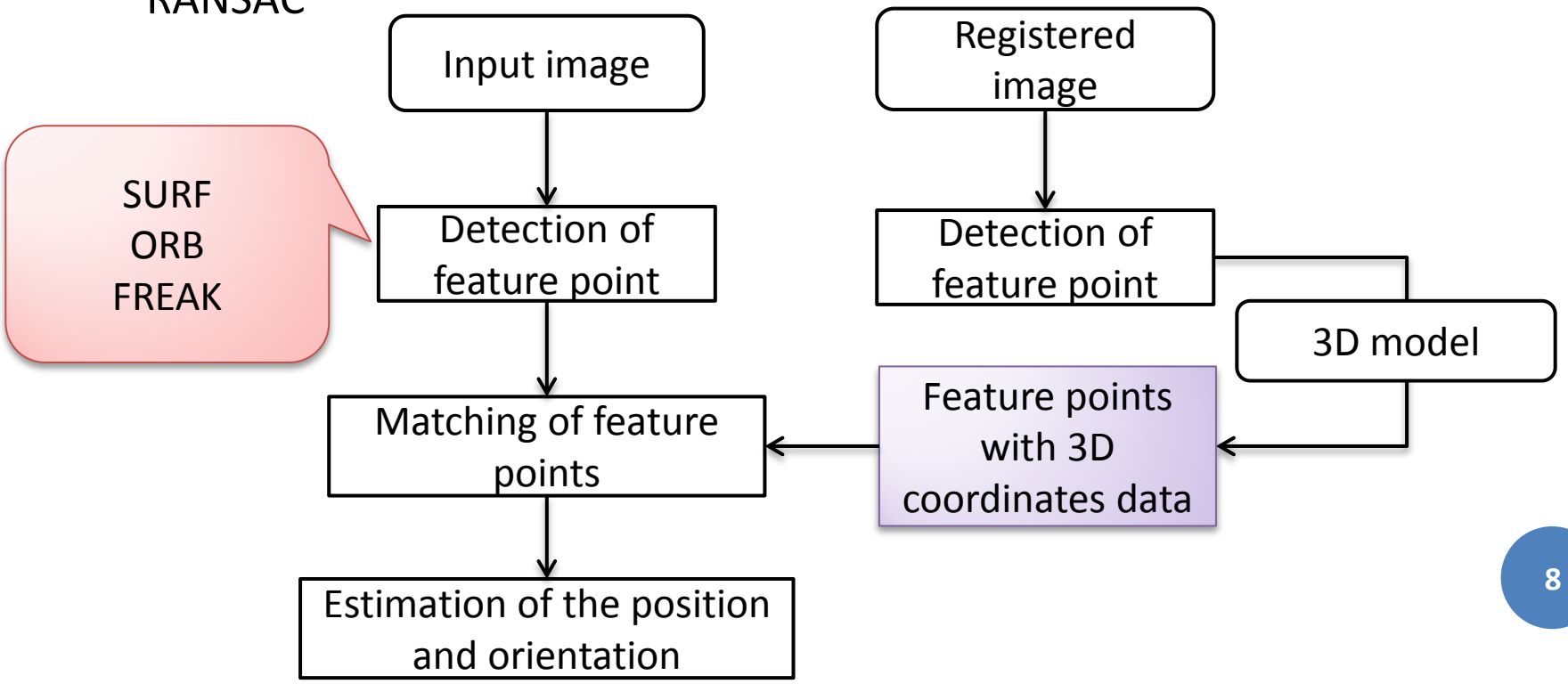
The registered image contain photo-shoot position, orientation, and depth information.



The method using feature points

Processing steps

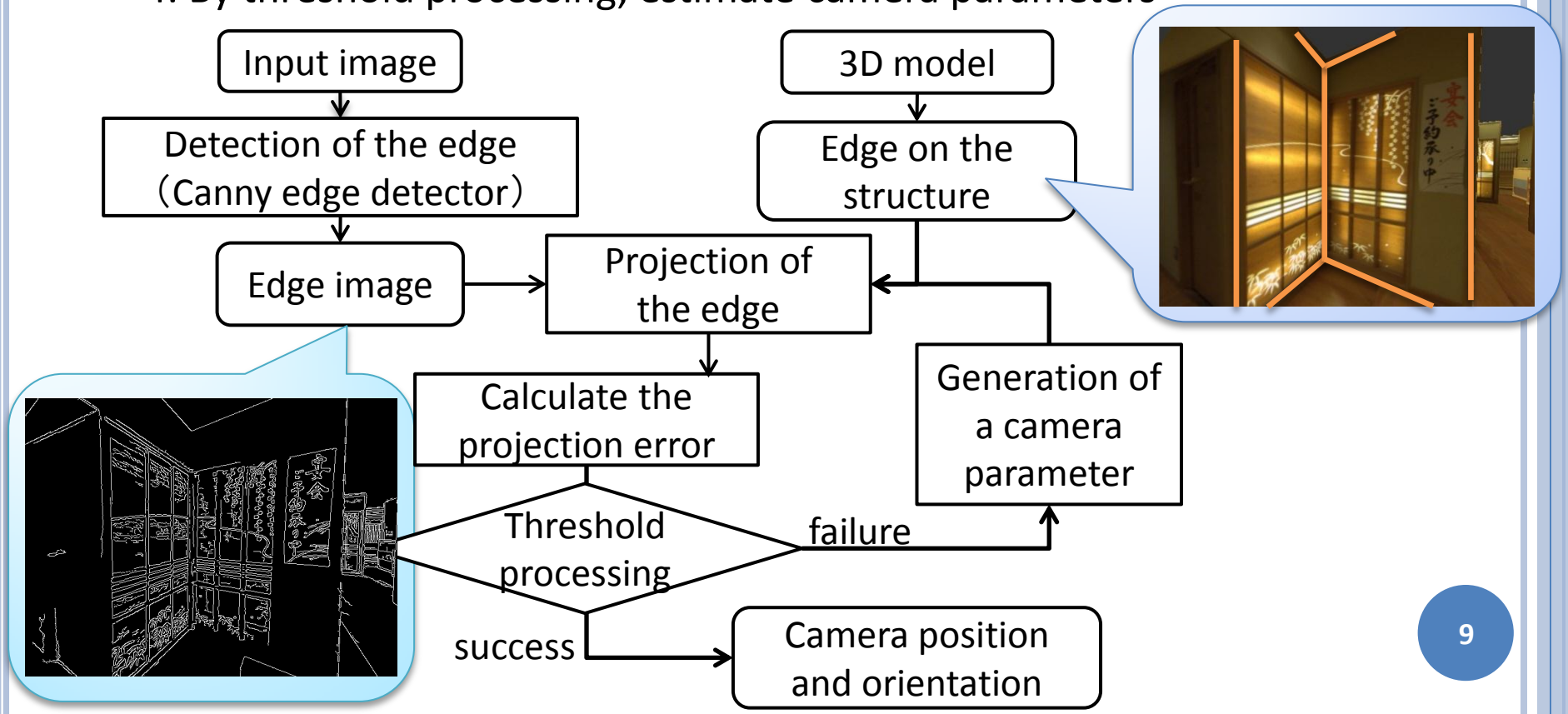
1. Detection of feature points from the input image and the registered image
2. Matching of feature points
3. Estimation of camera parameters from a set of feature points by RANSAC



The method using edge information

Processing steps

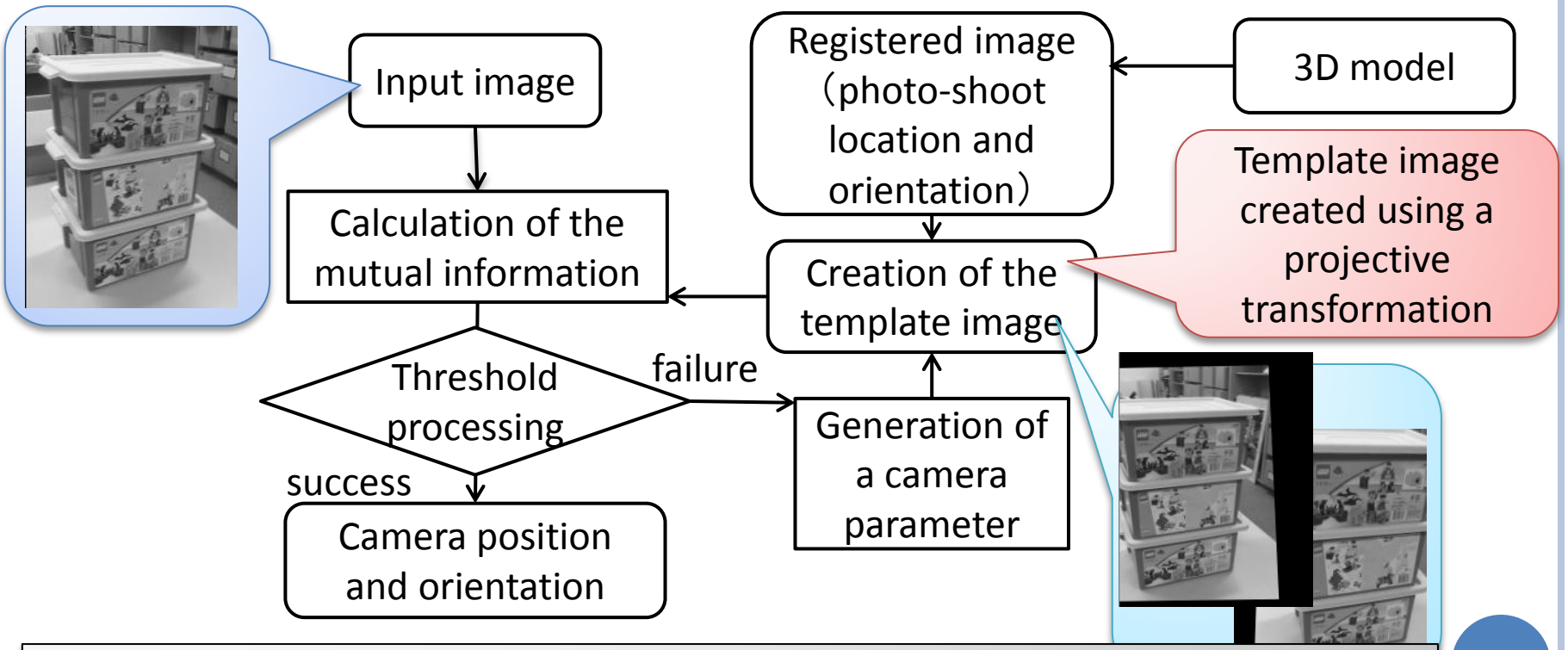
1. Detection of the edge (Canny)
2. Projection of the edge from the structure using generated camera parameters
3. Calculation of the projection error of the edge
4. By threshold processing, estimate camera parameters



The method using mutual information

Processing steps

1. Creation of the template image
2. Calculation of the mutual information
3. Estimation of camera parameters by threshold processing



A. Dame, E. Marchand. "Accurate real-time tracking using mutual information, "Proc. IEEE International Symposium on Mixed and Augmented Reality (ISMAR2010), pp. 47-56, 2010.

Accuracy evaluation for combination with the relative positioning

We apply following indexes for accuracy evaluation.

For the standardization, we have been discussing about the indexes in TrakMark

- Position error
- Orientation error
- Projection error
- Processing time

- Position and orientation error
  - For correction of relative positioning, it is important that it can be judged as failure if error is large
  - Threshold processing using confidence can be applied for the judgment
- Processing time
  - Affect the delay of the correction of the relative positioning
  - Estimated result can be used if there is a delay

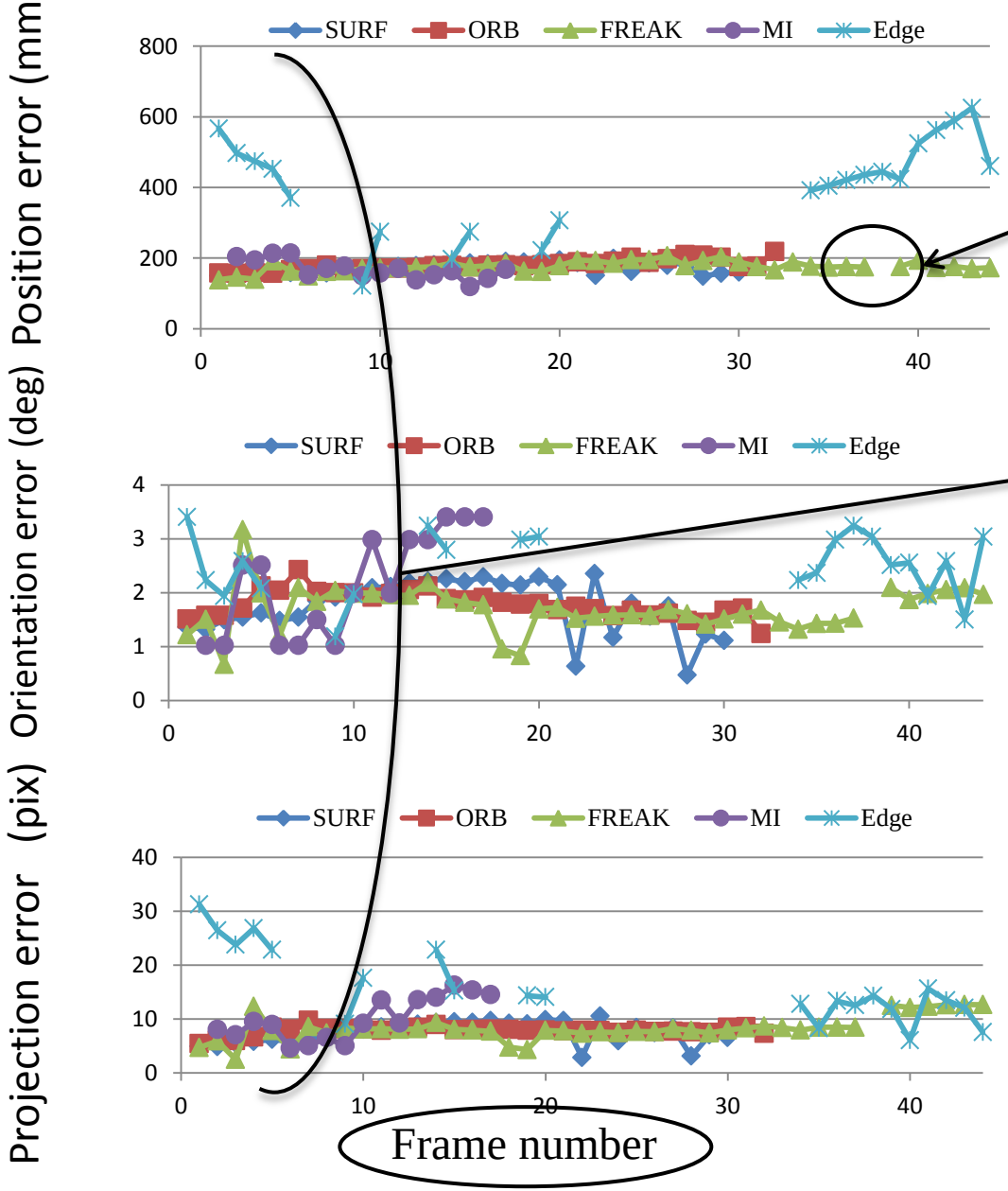
K. Makita, et al: “ A benchmark suite with virtualized reality models for supporting evaluation and data set generation, ”In Proc. of TrakMark2012, 2012.

Input images ～scene 1/3（640*480）～

In virtual environment

Both input images and registered images are CG (Generated from 3D model).

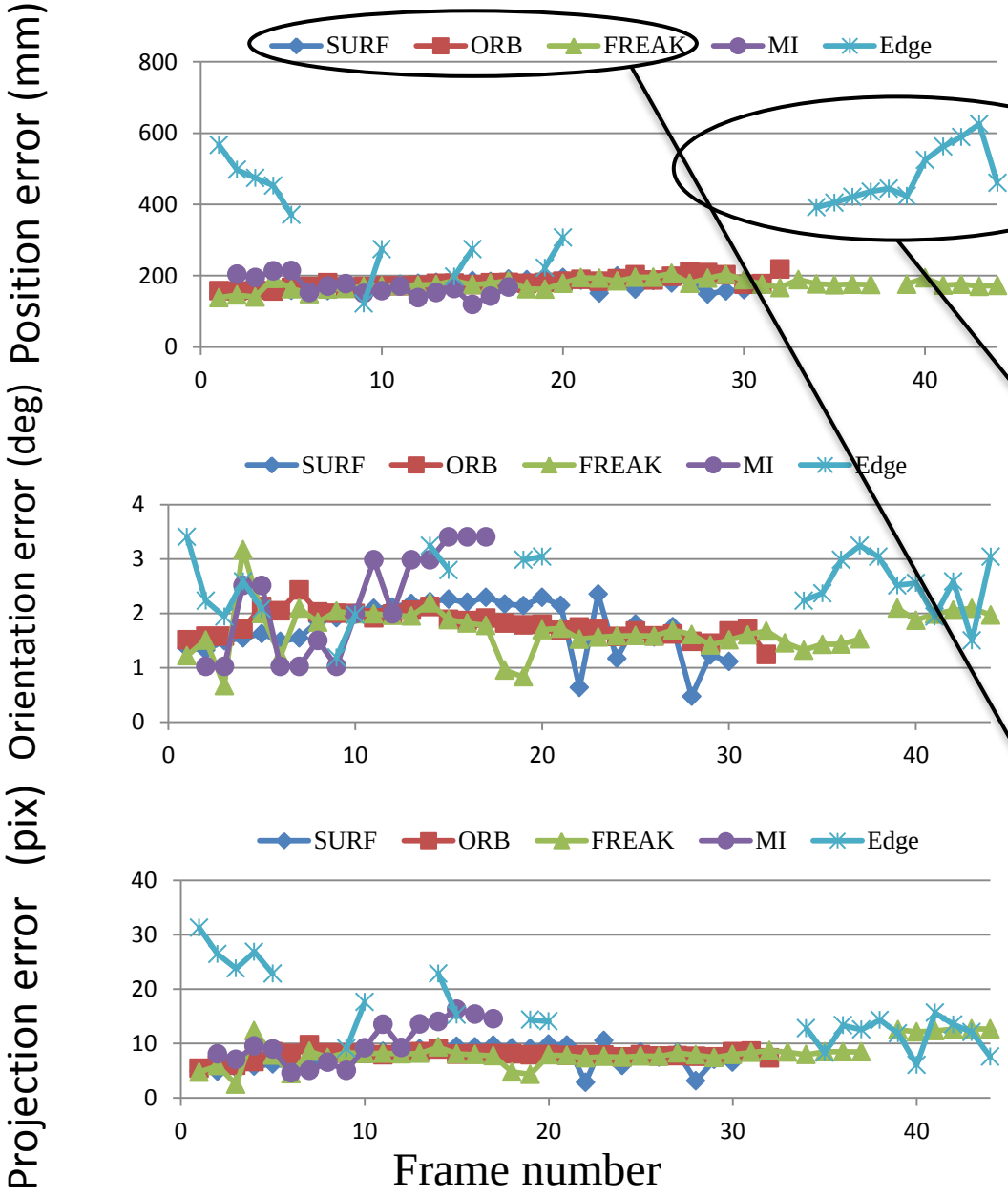
Description of the graph



No plot =
Estimation is failed

Error

Result ～scene1～



～Virtual environment～
Best scene calculating the correspondence of features

Despite position error is large, the method using edge information judged the result as success

Results using feature points are good in many frames.
（SURF、ORB、FREAK）

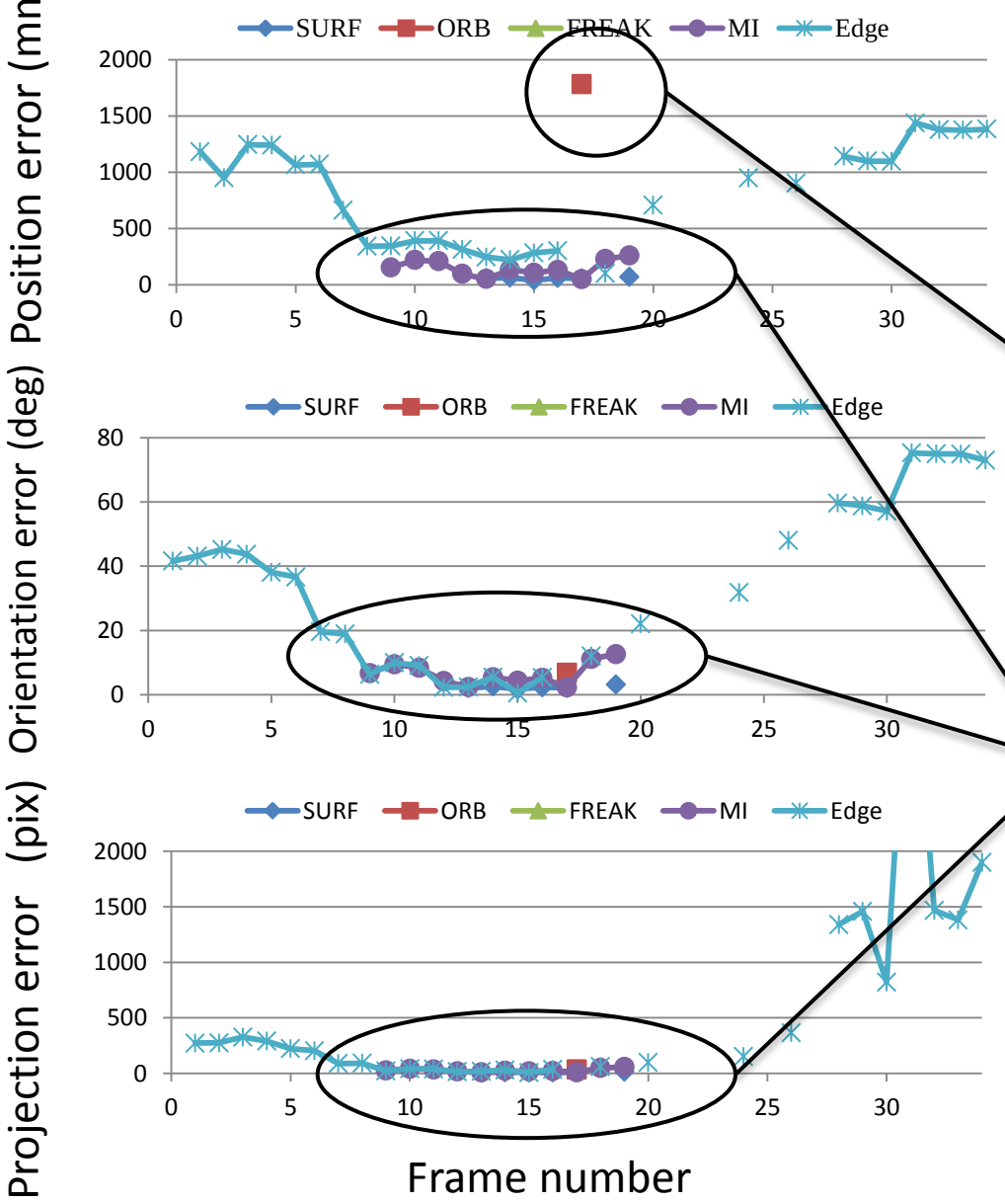
Input images ～scene 2/3（180*240）～

In real environment

A target object has complex textures, which is supposed to be better for methods using feature points.

A resolution of images are low of send image through the network.

Result ～scene2～



Real environment with complex textures

Results using ORB included large position error in the front-back direction.

Results using Mutual information (MI) and SURF were good, and judged as success.

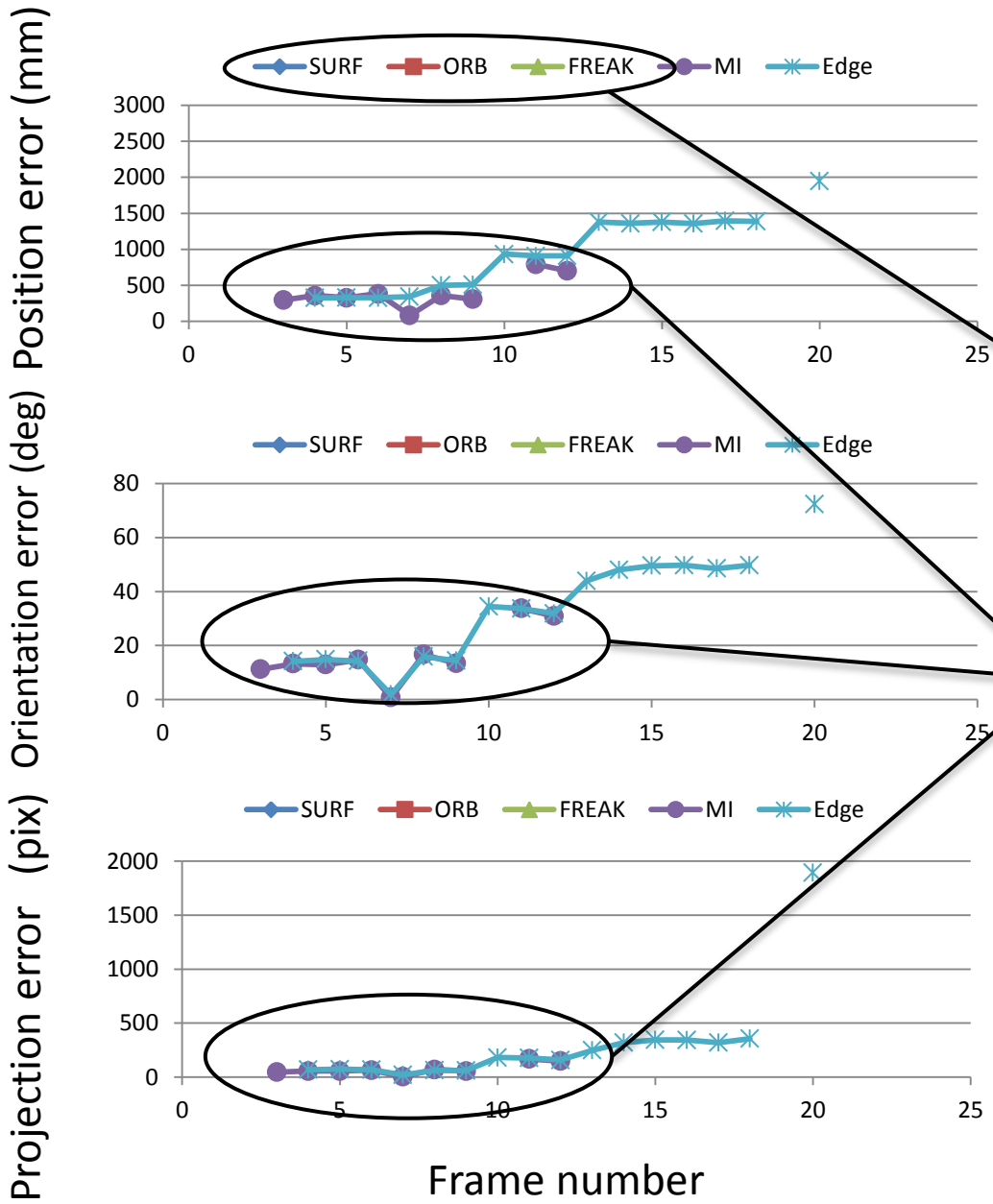
Input image ～scene 3/3（180*240）～

In the real environment

There are little complex textures, which is supposed to be better for methods using edges and mutual information.

A resolution of images are low of send image through the network.

Result ～scene3～

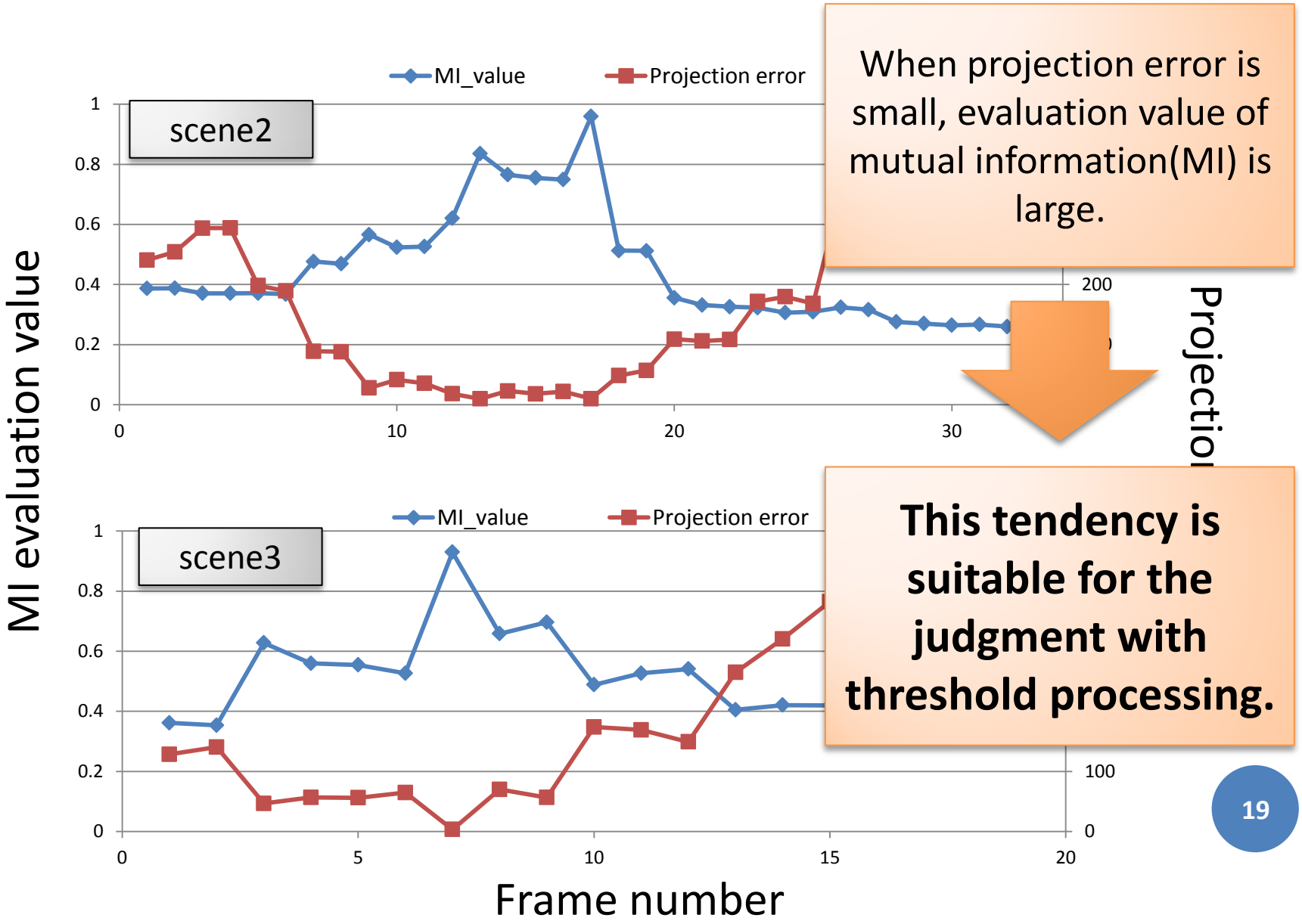


Real environment with less complex textures

Results using feature points were failed in all frames.

Results using mutual information was best.

Comparison between an evaluation value of mutual information and projection errors



When projection error is small, evaluation value of mutual information(MI) is large.

This tendency is suitable for the judgment with threshold processing.

Processing time of the localization

In methods using feature points (ms)

	SURF	ORB	FREAK
Resolution of input image (180*240)	300.2	339	305.3

In method using edge and mutual information (ms)

	Edge	MI
Resolution of input	79.8 × N	27.7 × N

N is number of the template

- Each number indicates a time for generating one template and calculation
- Required processing time proportional to the number N
- In the experiments, 3125 templates (125 positions * 25 orientations) were used

Mutual information(MI) was highly accurate and processing time was to evaluate and short with one camera parameter. But spending up is necessary for actual use.

Summary

- We propose a method to estimate the position and orientation of the camera with real image and registered image in 3D model that assumes the cooperation with relative positioning

- Experiments result in 3 scene
  - In the method using feature points, SURF method was small error.
  - Mutual information(MI) was high accuracy to evaluate and short processing time in one camera parameter. But need to increase the number of templates to extend the search area.
    Therefore, it is necessary to devise faster for actual use.

	scene2		**scene3**	
	SURF	MI	SURF	MI
Accuracy	○	○	×	○
Time	○	△	○	△

Future work

- Edge detection from registered image
  - Currently available only on the edge structure
- Parallel processing in mutual information method
- By the combined use of the method using the mutual information and using SURF, expansion of the area can be estimated.
- Acceleration method used in the database search
- Estimated by the interaction with the user assuming the AR and navigation applications
  - Presents the estimated location and direction will be easy to succeed. Get the image to facilitate the estimation by cooperation of the user. (推定が成功しやすいであろう場所、方向などを提示し、ユーザと協力して、推定が容易な画像を取得)