RGB-D Model Based Human Detection and Tracking Using 3D CCTV [1)]

Seohee Park, Junchul Chun

Department of Computer Science, Kyonggi University
Suwon, South Korea

1) This research was supported by Basic Science Research program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (No 2018R1D1A1B07042498).

Contents

1. Introduction

- Nowdays, the basic CCTV system has been combined with various fields such as big data, image analysis, and artificial intelligence.
- It has evolved into an intelligent CCTV that can detect and analyze the overall situation of objects such as pedestrians.
- A variety of image analysis researches has been conducted to recognize such situations as crime, fire, automatic incident detection, and anomalies using such intelligent CCTV.

Type	Feature	Example
Basic CCTV	People monitoring video	Video surveillance 
Intelligent CCTV	Automatically identify people and things	Video analysis device 

▲ Basic CCTV vs Intelligent CCTV

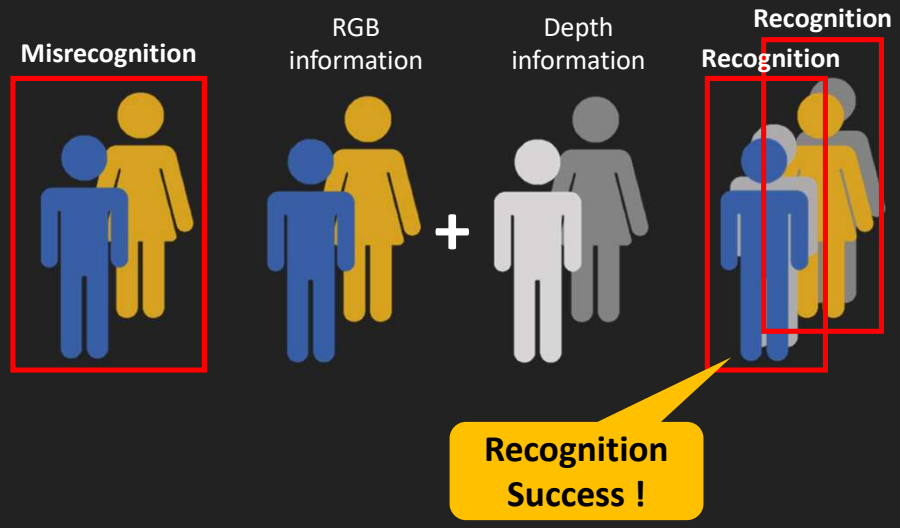
1. Introduction

- However, the CCTV images using two-dimensional information generally lack the topological information.
- In order to solve the object misrecognition problem caused by the loss of information by projecting the 3D real world into 2D image, the object can be more correctly recognized by combining the depth information with Color information.

▲ A problem of object misrecognition

Research Goal

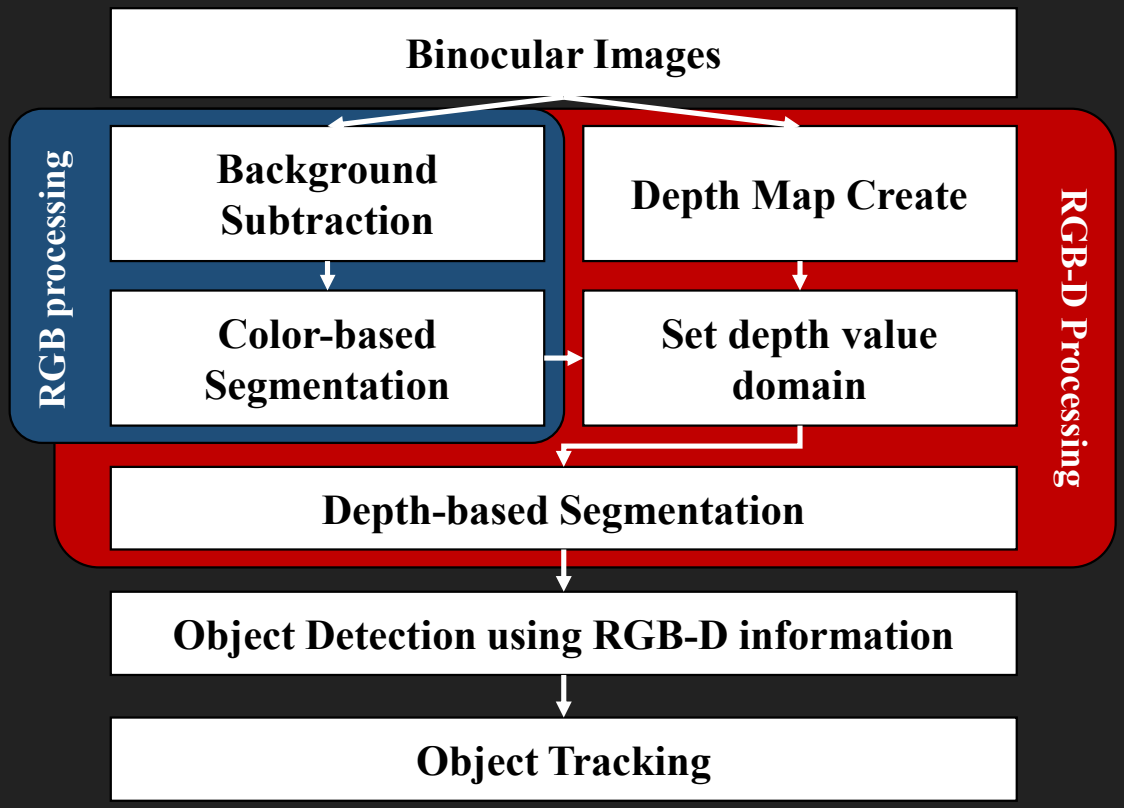
This paper proposes an method to detect and track moving objects by combining depth information with RGB information in two CCTV environments.



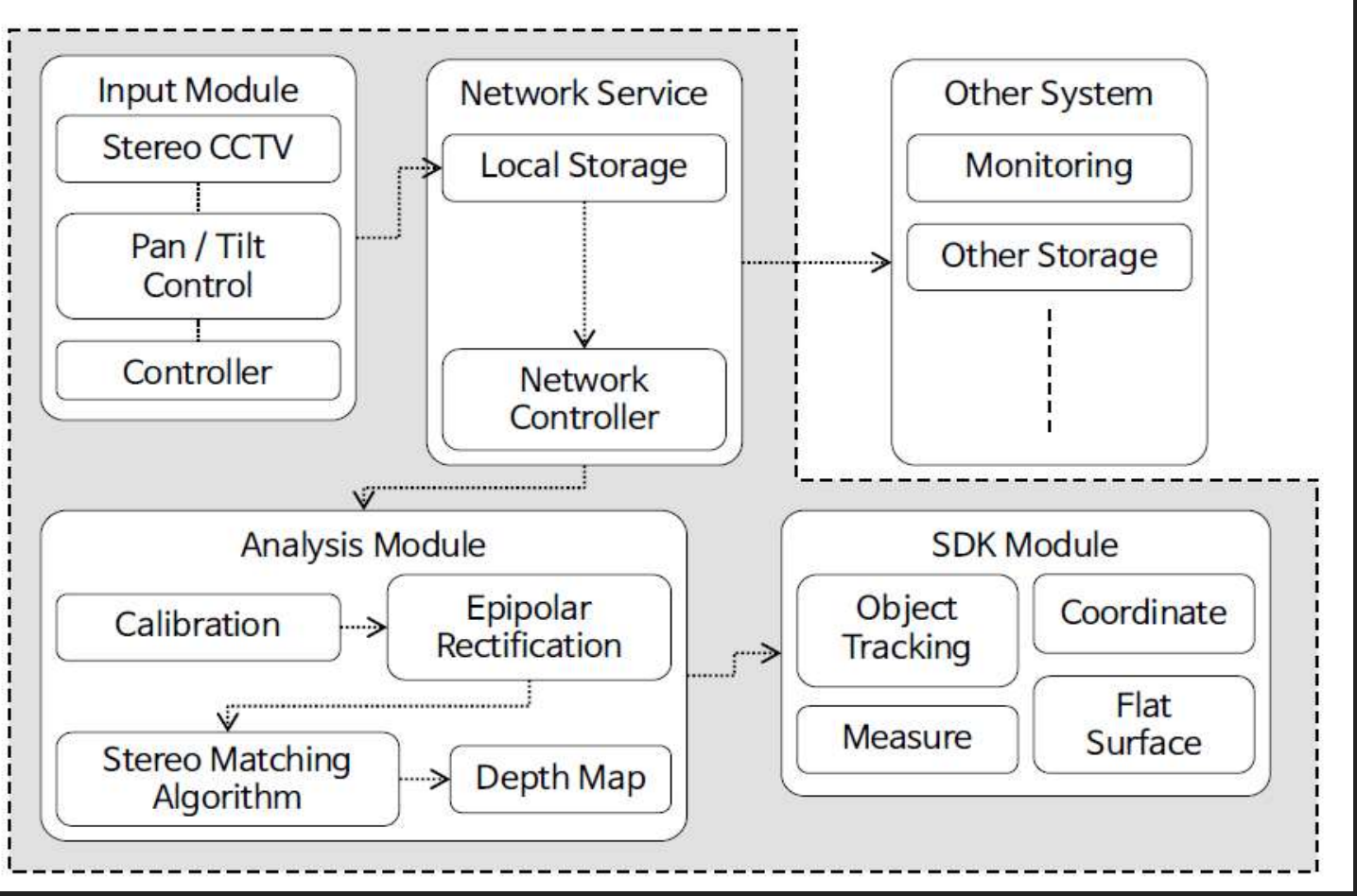
2. Object Detection and Tracking using RGB-D information

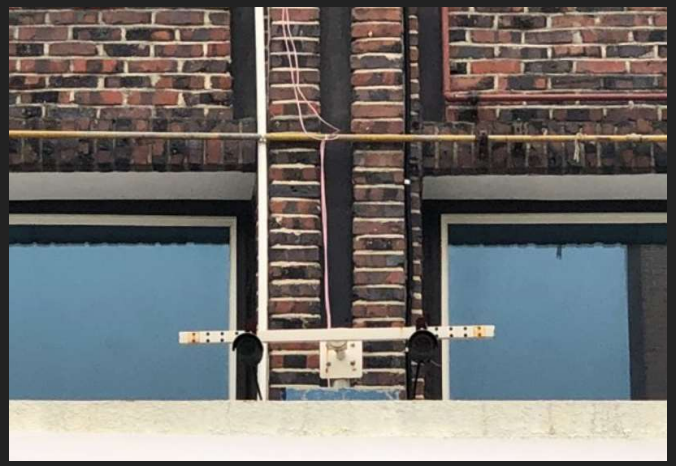
3D CCTV based object detection and tracking algorithm using RGB-D information

- Object Detection and Tracking using RGB-D information
  - Color-based Segmentation
  - Depth-based Segmentation
  - Object Tracking
- Experimental Results
- Conclusions



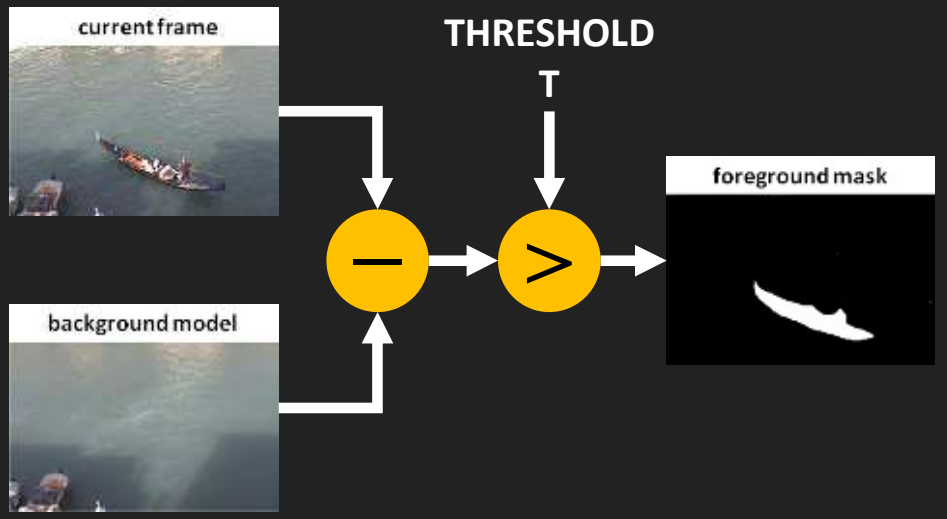
The Configuration of 3D CCTV System





1) Color-based Segmentation

- In an intelligent video surveillance system, moving objects must be robustly detected in order to track the behavior of objects.
- Background modeling is performed using Mixture of Gaussians technique, performed by calculating the initial background model and updating it continuously.
- The object is segmented from the background by performing a difference operation with the moving object in the modeled background.



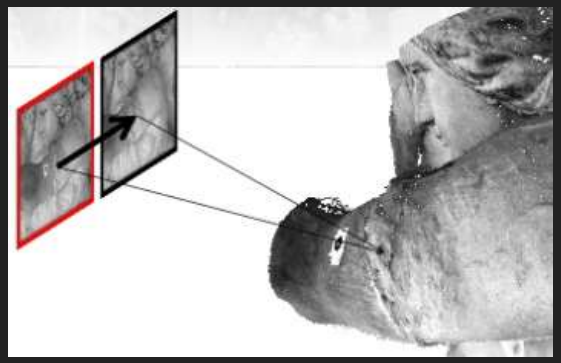
▲ Background modeling process

(a) Frame (b) Background Modeling (c) Morphology (d) Binarization

2) Depth-based Segmentation

- In order to obtain depth information, depth map is generated using two connected CCTV images.
- Generate depth map through disparity calculation.
  - The difference between the images input from the left and right cameras is called Disparity, which plays an important role in generating the depth map.
- In order to calculate the disparity, the similarity of the left and right images is measured, matched on a block-by-block basis using it.

(a) Left Frame (b) Right Frame (c) Depth Map

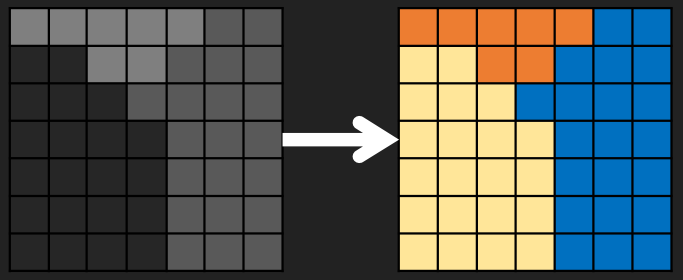


▲ Image Matching

2) Depth-based Segmentation

- In order to perform the segmentation based on the generated depth map, the result segmented by RGB is set as the domain region to extract the depth value.
- The depth value is sequentially searched in the set domain and the previous pixel value is compared with the current pixel value.
- Even if the object is the same object, since the depth value of the moving object may be slightly different, object segmentation is performed according to a certain depth value range.

(a) Frame (b) RGB –based segmentation (c) Depth Map (d) Domain setting



▲ Image Matching ▲ Depth based Segmentation

2) Depth-based Segmentation

- The case of being included in a certain range value is regarded as the same region, and the case of not including the range is regarded as another region and the segmentation is performed.

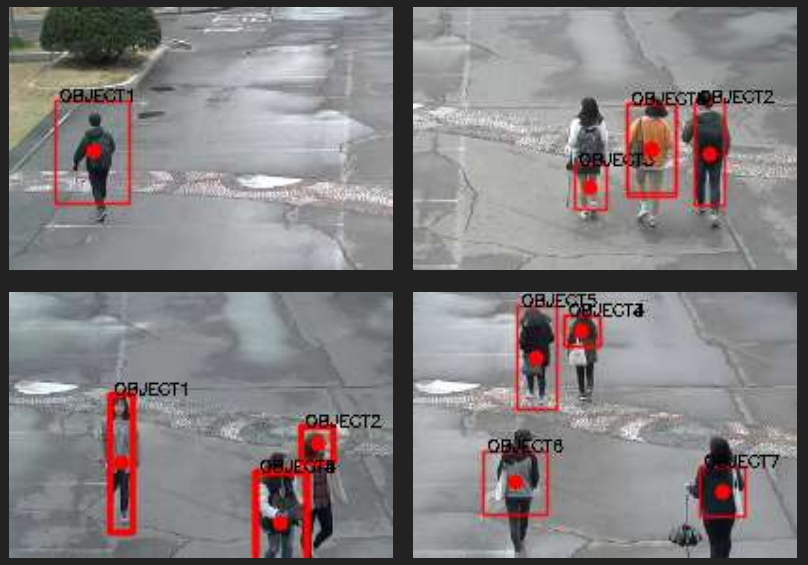
$$-\alpha < Depth\ Value < +\alpha$$

- Advantage
  - By setting the domain, it is possible to reduce the amount of computation than searching the depth value in the all area.
  - Because the disparity is clear when the object is moving, we can use only the appropriate depth value that does not include noise.

(a) Frame (b) RGB-based Segmentation (c) Depth-based Segmentation (d) Object Detection

3) Object Tracking

- Object Tracking is an essential process of an intelligent video surveillance system that obtains information such as the moving direction of objects, center point of objects.
- The motion of moving objects is tracked by applying CAMShift(Continuously Adaptive Mean Shift) algorithm.
- The CAMShift tracking method moves along the average value of the data and finds the peak or center of gravity of the distribution.
- The color histogram of the object to be tracked is compared with the histogram of the current input image to find the most similar window region.



▲ The result of tracking with CAMShift using RGB-D model

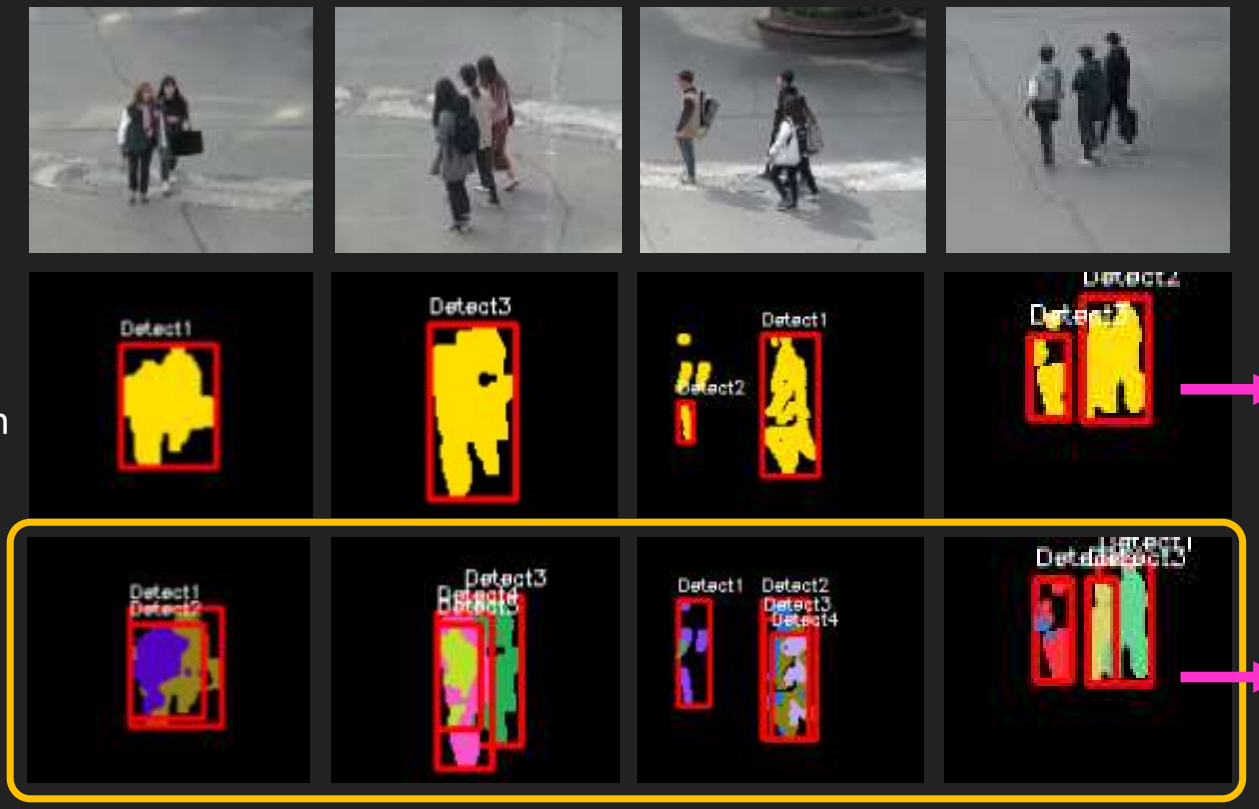
3. Experimental Results

- The comparative results between object detection using only RGB information and RGB-D information.

(a) Frame

(b) Object Detection using only RGB information

(c) Object Detection using RGB-D information



Number of detected object = 2

Number of detected object = 3

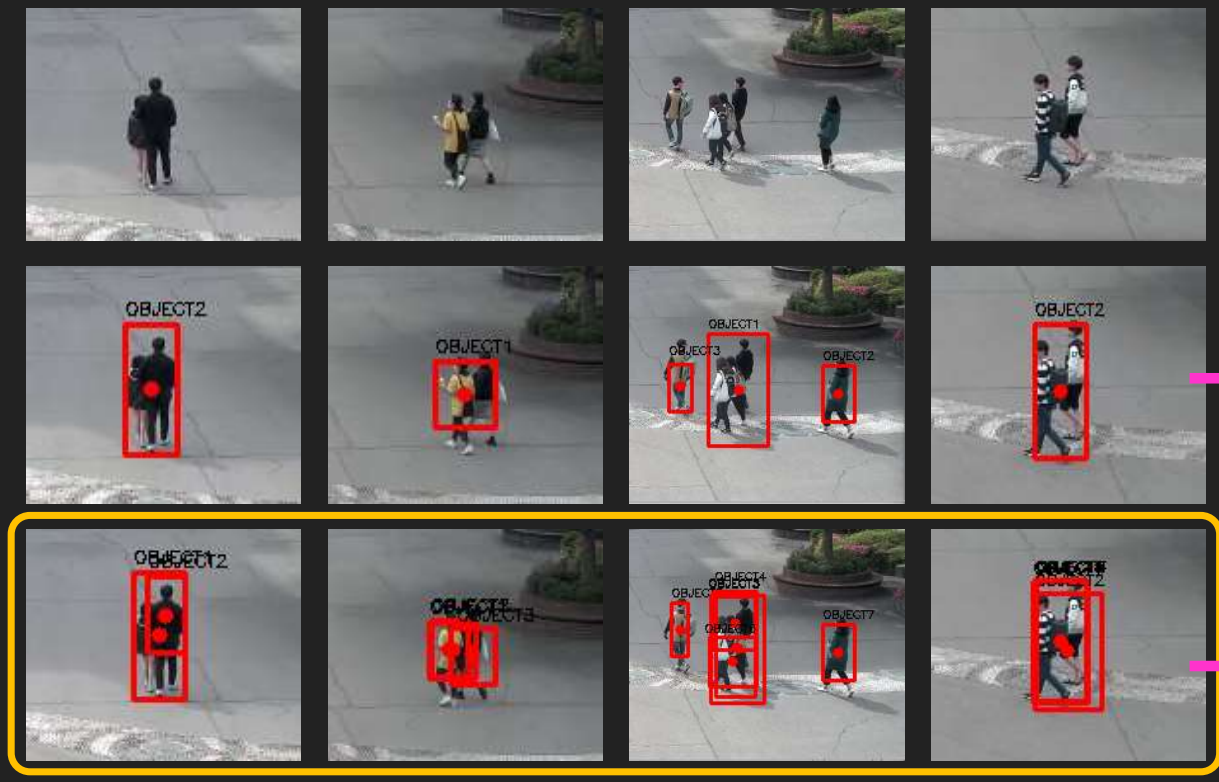
3. Experimental Results

- The comparative results between object tracking using only RGB information and RGB-D information.

(a) Frame

(b) Object Tracking using only RGB information

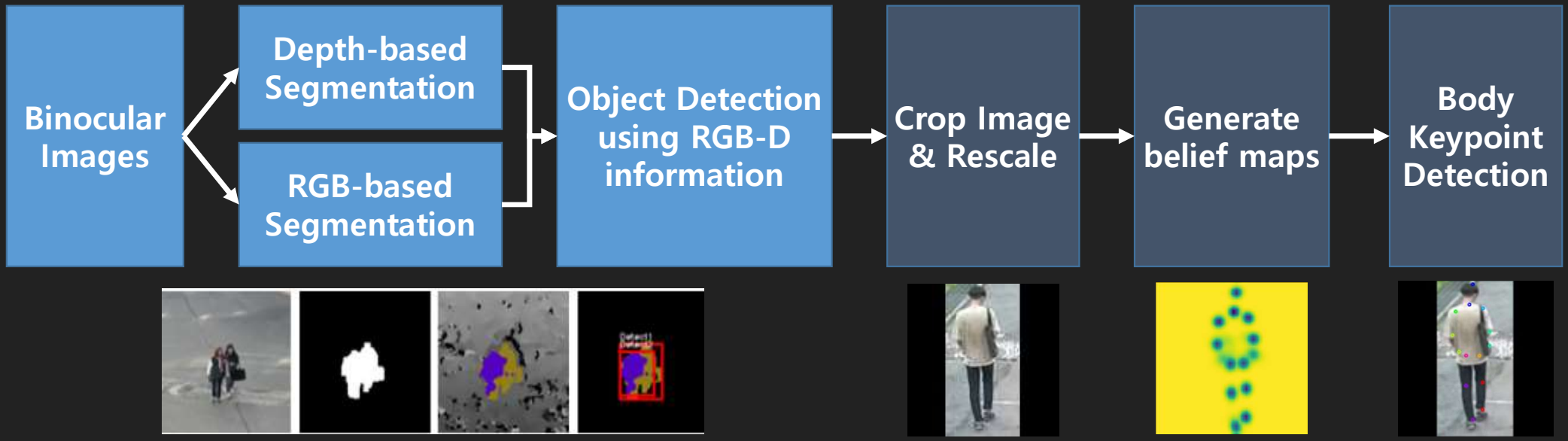
(c) Object Tracking using RGB-D information



4. Conclusions

- In this paper, 3D CCTV based object detection and tracking using RGB-D information is performed.
- As a result, the problem of object misrecognition when segmenting an object using only RGB information is improved by adding depth information.
- In future research, the object detection and tracking method based on the RGB-D model proposed in this paper can be applied to expand the scope of the research into the field of human activity recognition.

5. More Works


Binocular Images
Depth-based Segmentation
RGB-based Segmentation
Object Detection using RGB-D information
Crop Image & Rescale
Generate belief maps
Body Keypoint Detection

References

1) Grant, Jason M., and Patrick J. Flynn., "Crowd Scene Understanding from Video: A Survey," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, Vol 13, No. 2, pp. 19, 2017.

2) Zivkovic, Zoran, "Improved adaptive Gaussian mixture model for background subtraction," *Proceedings of the 17th International Conference on Pattern Recognition*, pp. 28-31, 2004.

3) Hirschmuller, Heiko, "Stereo processing by semiglobal matching and mutual information," *IEEE Transactions on pattern analysis and machine intelligence, Vol* 30, No. 2, pp. 328-341, 2008.

4) Bradski, Gary R, "Computer vision face tracking for use in a perceptual user interface," *Intel Technology Journal*, pp. 214-219, 1998.

THANK YOU FOR YOUR ATTENTION !

If you have any further questions, please email me.

jcchun@kgu.ac.kr