CCTV based Video Analytics for Managing and Forecasting Impact of Weather

Hanseok Ko

School of Electrical Engineering
Big Data & AI Institute
Korea University, Seoul

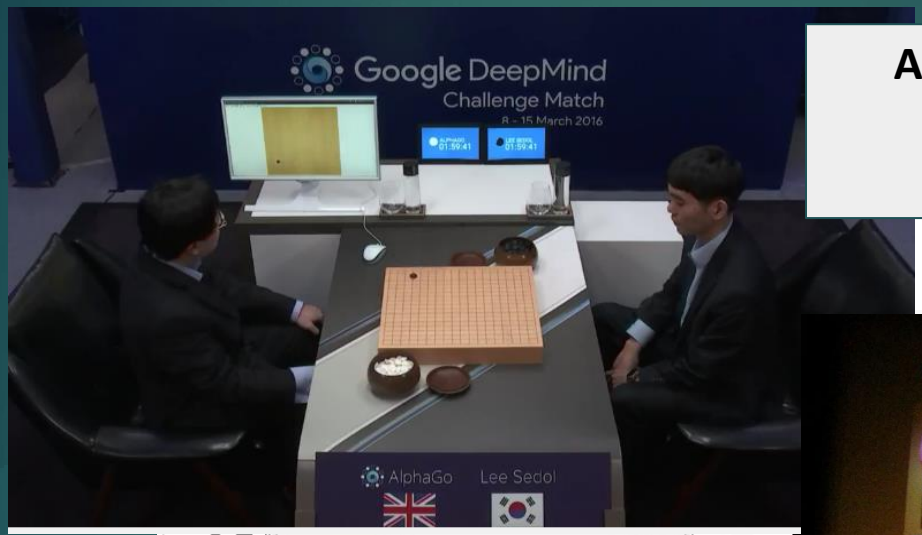
2016.01.19~22 DAVOS FORUM

4th Industrial Revolution

- AI
- Big Data Analytics

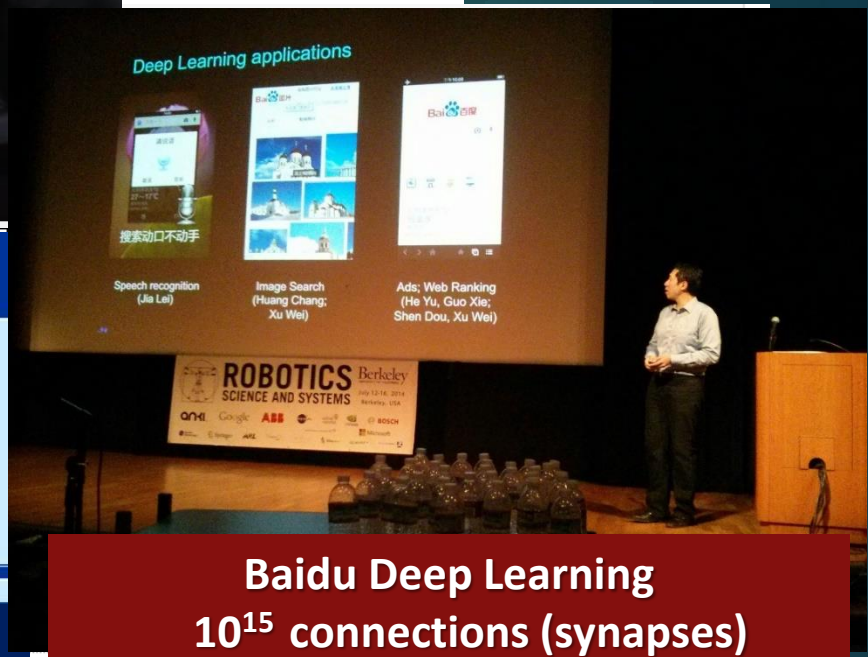
Chat-bot, Diagnostic, Weather Prediction

Machine that can think like human



Alpha-Go vs. Sedoll LEE
4 vs. 1
2016.03

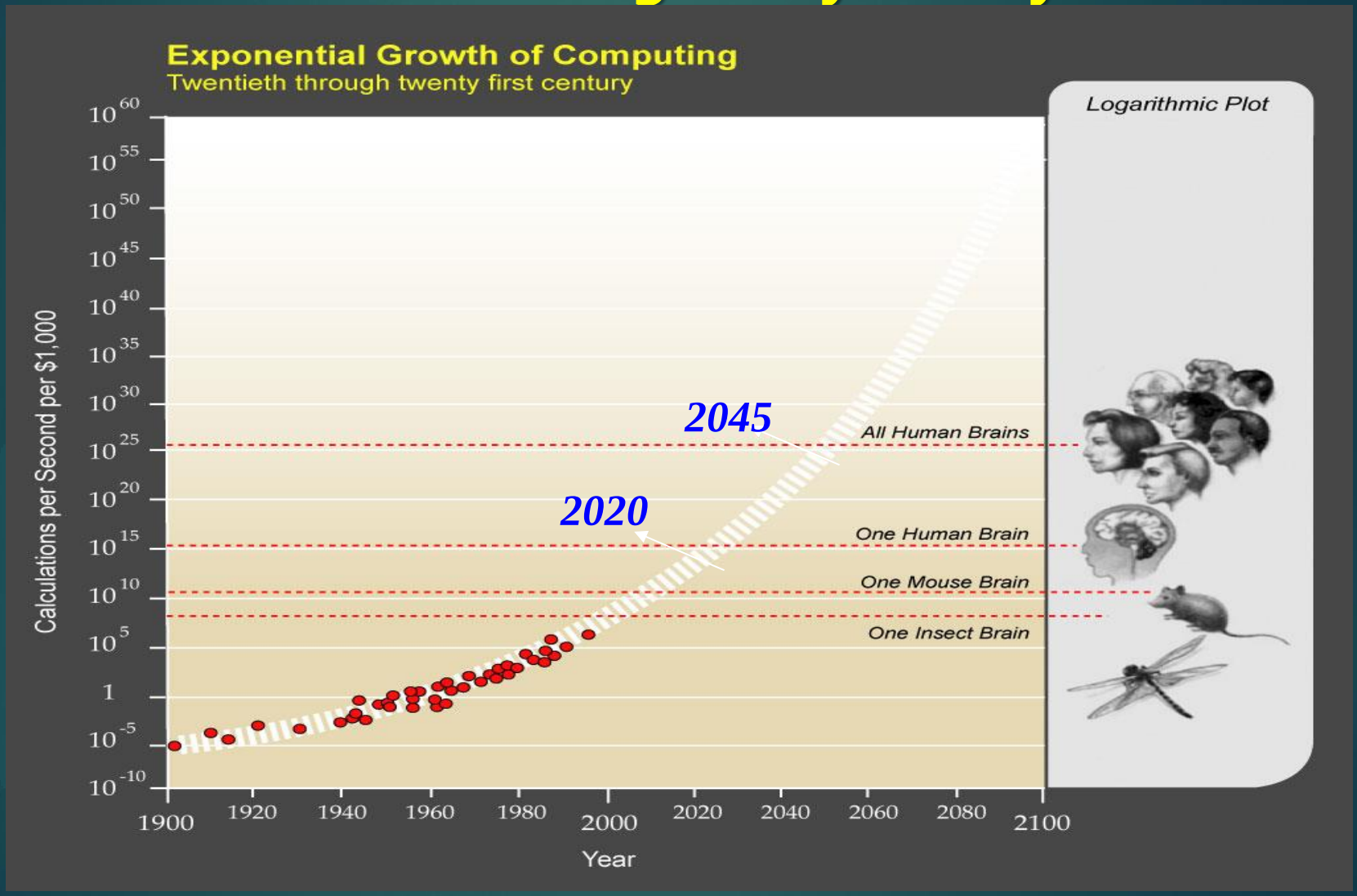
Computer program convinced humans as
a 13-year-old boy to pass the Turing Test
2014



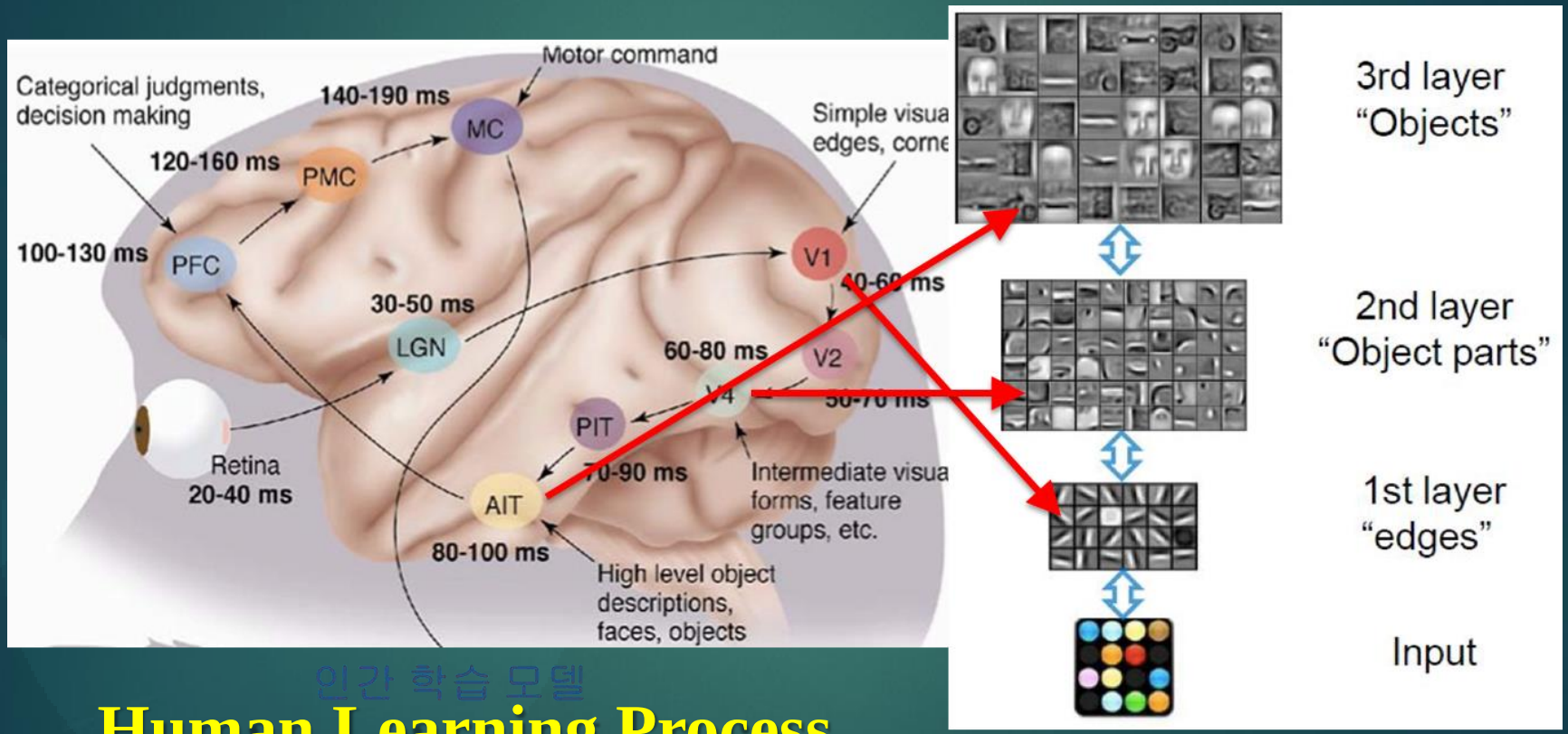
Baidu Deep Learning
10^{15} connections (synapses)
neural connection system
2015

Accelerating Intelligence :

"Singularity" – Ray Kurzweill



Deep Neural Networks: human cognition?

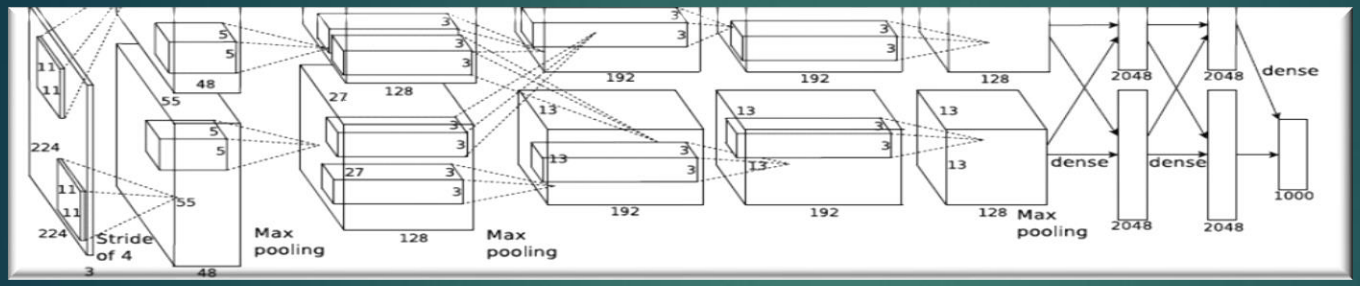


인간 학습 모델

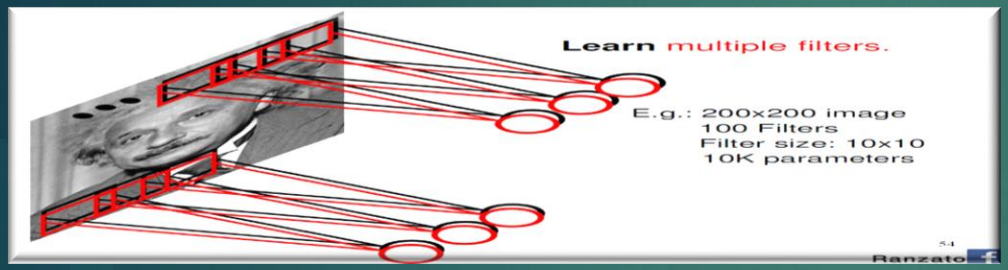
Human Learning Process

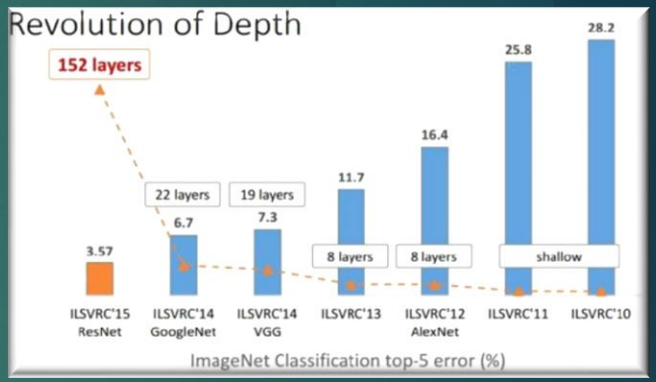
Deep Learning

Deep Neural Networks



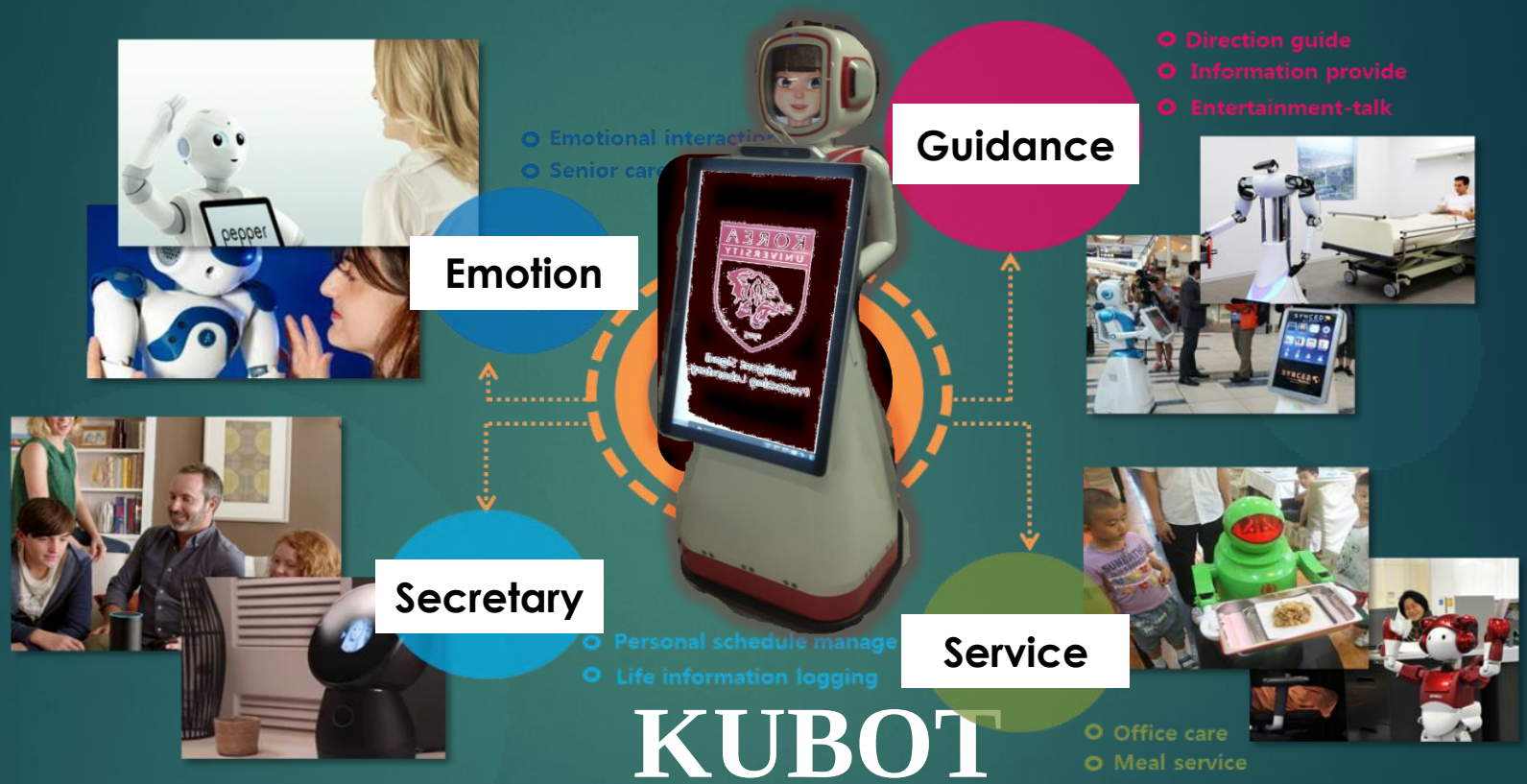
Alexnet





Inception Networks (GoogleNet)

AI / Machine Learning based Vision /Speech Understanding



: Speech interfaced conversational robot

Speech enhancement / noise reduction

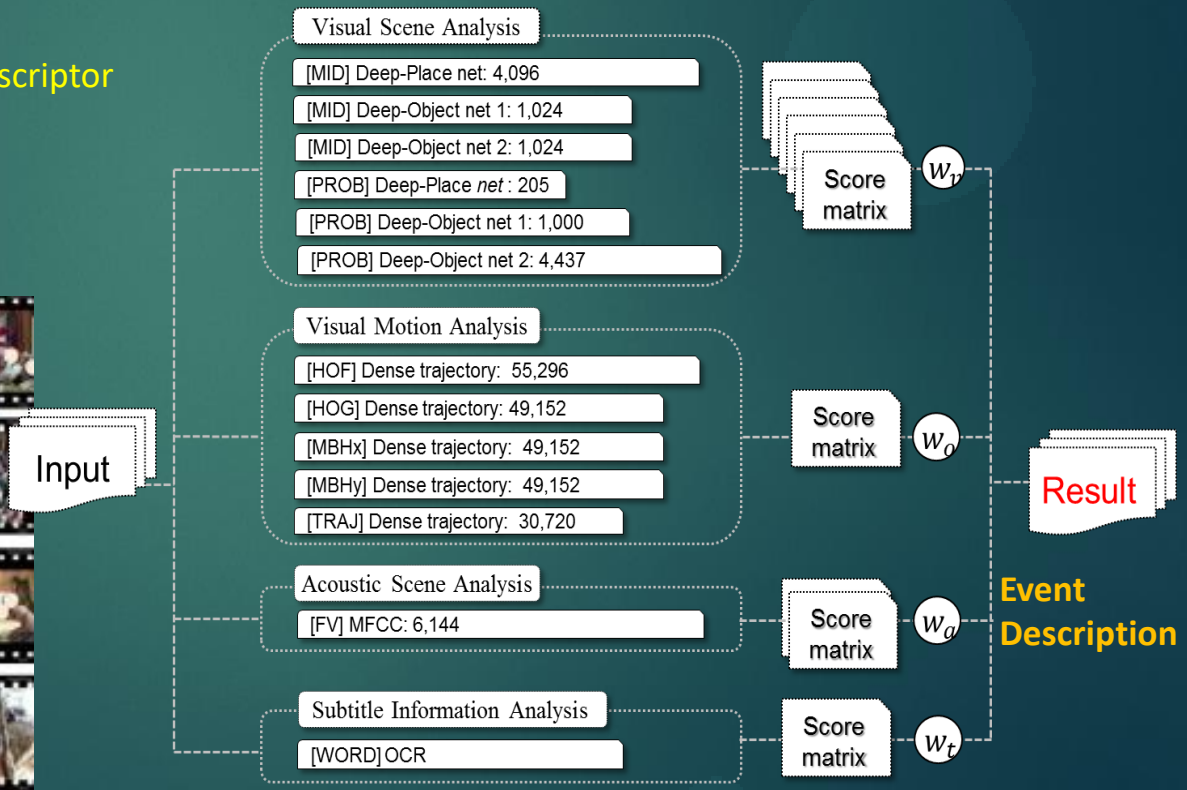
Speaker identification (speech/face)

Speech recognition based on Deep learning

Natural language processing

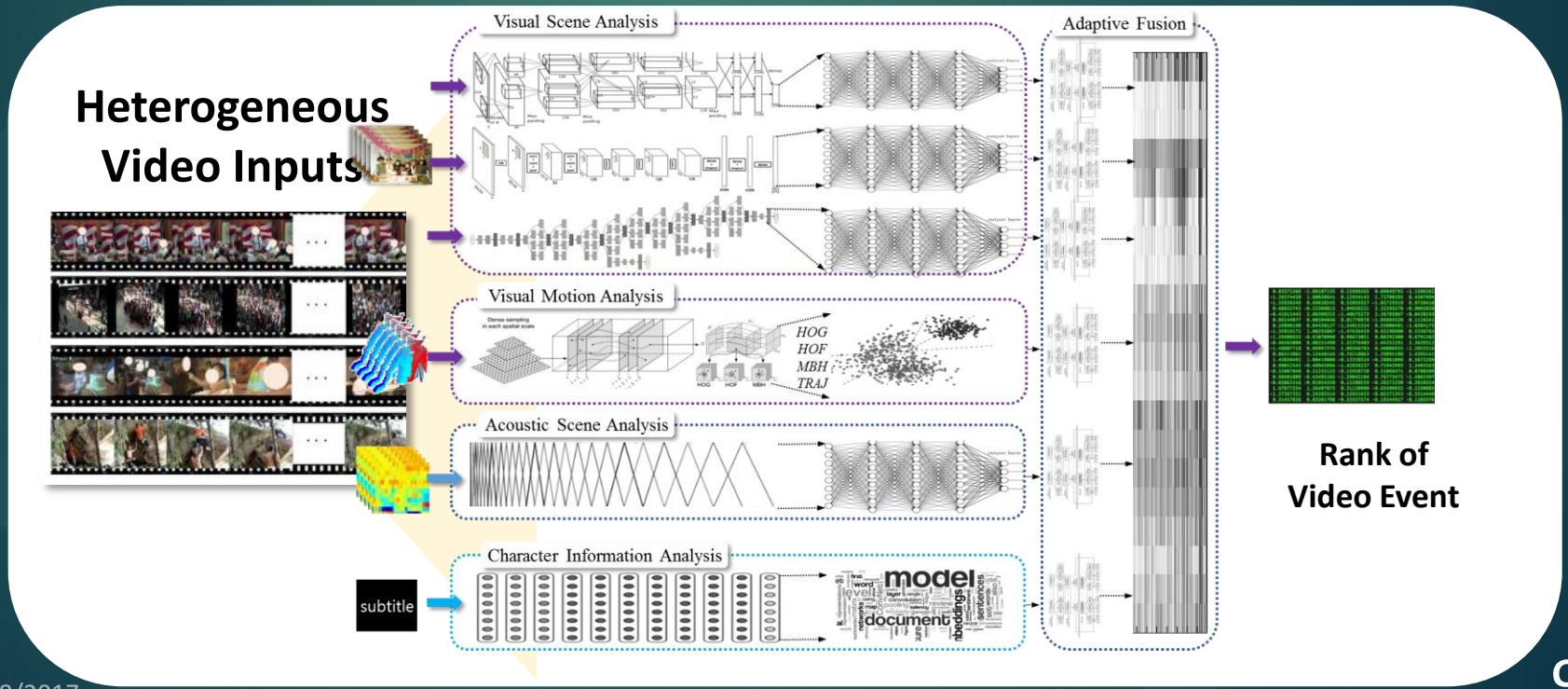
Big data Video Analysis

- **Visual scene analysis**
  - Visual Motion: HOG, HOF, SIFT, Dense trajectories
  - Visual Scene: Deep-CNN based object/scene descriptors
- **Acoustic scene analysis**
  - Low level: MFCC, TM
  - Semantic Level: TF-IDF based descriptor
- **Speech recognition**
- **Optical character recognition**



TRECVID
– Global Challenge for video analytics

- Event Query Generation (modeling)
  - Deep architecture-based hybrid video event learning
- Event Search (detection)
  - Two-fold information fusion
    - Dynamic Feature Selection Adaptive Metadata Weighting



TRECVID
– Global Challenge for video analytics

- **Training / Evaluation database (2016's)**
  - HAVIC (*NIST*), YFCC100M (*Yahoo Corp.*)
    - Heterogeneous Audio Visual Internet Collection (HAVIC) – *Default set*
      - User generated video collected from internet (***Youtube***)
      - Testing set : 100,000 video clips (EvalFull) / 16,000 (EvalSub)
      - Training set: 10 and 100 video clips for each video event (10EX / 100EX)
    - Yahoo Flickr Creative Commons 100M
      - User generated video collected from internet (***Flickr***)
        - Testing set : 100,000 video clips (EvalFull) / 16,000 (EvalSub)

An example of video event

1. ***Land Vehicle Accident event***
    1. **Definition**: A motorized land vehicle being operated by a human hits or crashes into another vehicle or object.
    2. The vehicle initiating the accident must be a motorized land vehicle
    3. **Evidential Description:**
        1. **Scene**: outdoors, roads or highways, person, …
        2. **Objects/People**: motorized land vehicle, car, …
        3. **Activities**: driving, loss of vehicular control, people observing
        4. **Audio**: vehicle impacting, engine noise, skid, …

Surveillance (Human Activity Recognition)

1. Event detection is the basic technology of intelligent surveillance system
2. Heterogeneous sensor fusion is required for effective event detection

Alarm triggered

Intelligent Security System (ISS)

Integrated Security Center

more accurate, more efficient

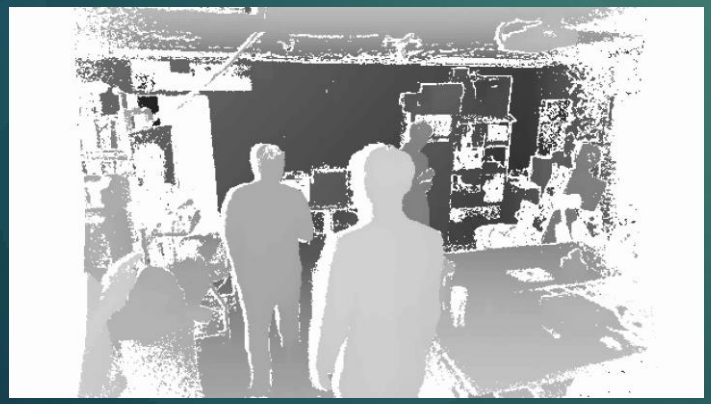
Surveillance

- **Feature extraction**
  - Interesting point extraction : Dense sampling
  - Point tracking : Optical flow tracking
  - Feature description : HOG (Histogram of Oriented Gradients) , HOF (Histogram of Optical Flow), MBH (Motion Boundary Histogram) , Trajectory
  - CNN

Interesting point tracking

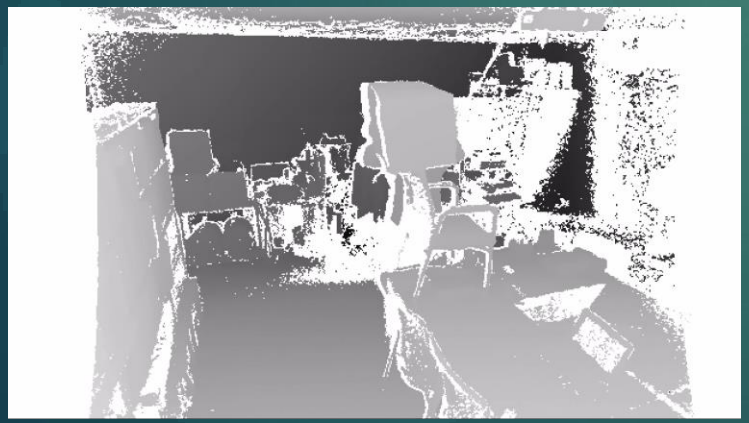
Surveillance

- Normal situation



Surveillance
- Demo

- Falling objects situation



Intelligent Transport System

5. Demo (Image Enhancement)

Intelligent Transport System

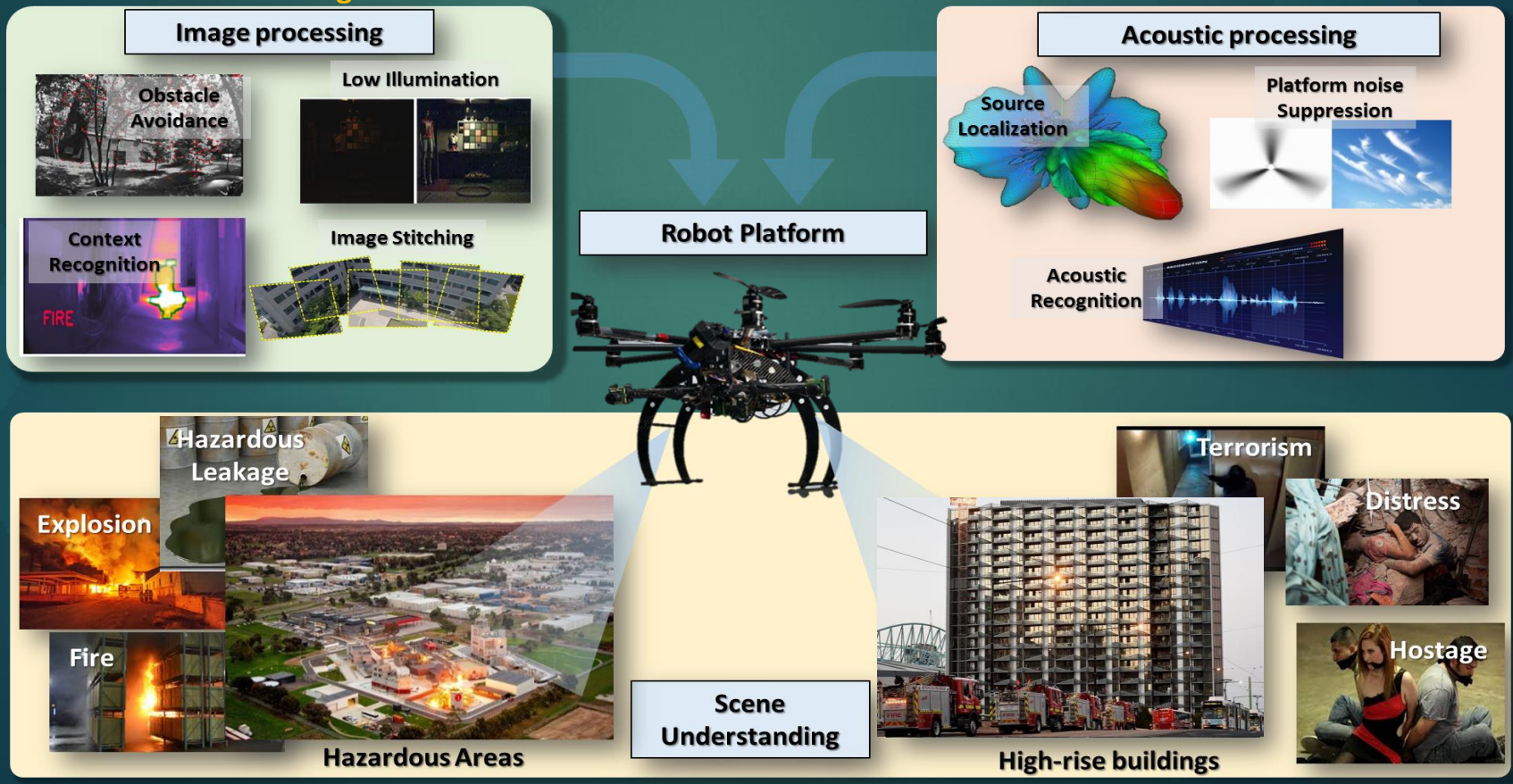
6. Demo (Image Enhancement & IR-CCD Fusion)

Hazy Video **Dehazed Video**

IR Video **CCD Video** **Fused Video**

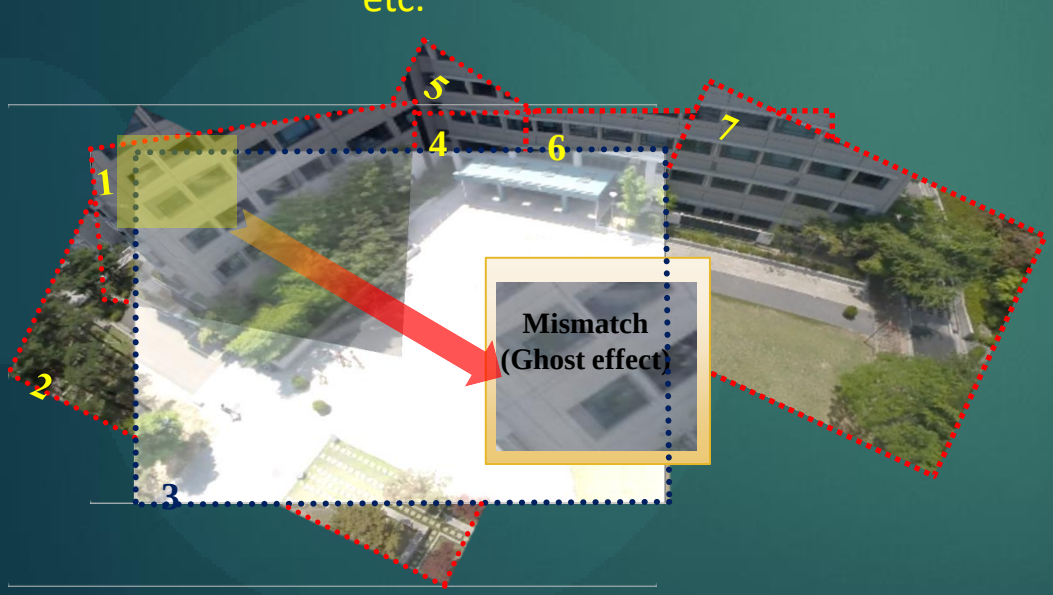
Drone for surveillance

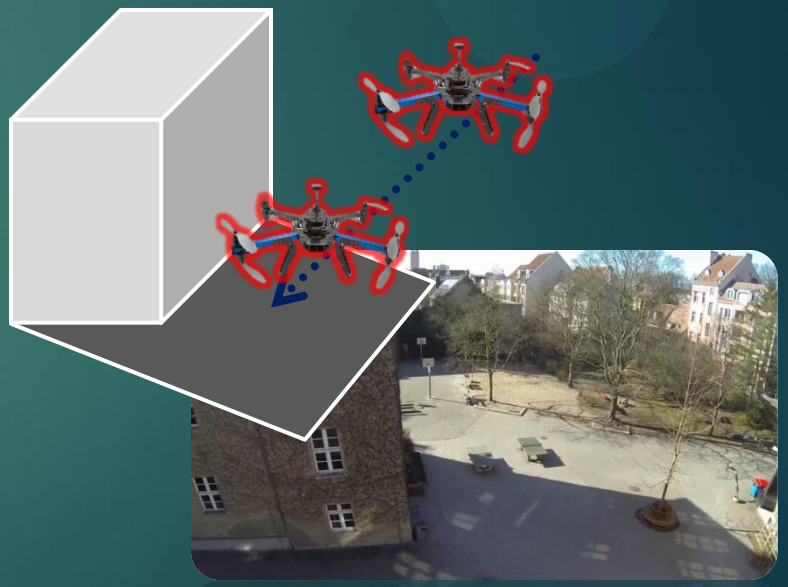
- **Image Processing**
  - Image Stitching, Obstacle Avoidance, Low Illumination
  - People detection / tracking, scene understanding
- **Acoustic Processing**
  - Platform noise suppression, Source Localization, Acoustic recognition
- **Scene Understanding**



Drone panoramic view system
- Challenges

- Poor performance in illumination changes
  - Image stitching has bad results in illumination changing situations.
  - Due to the characteristics of the drones, the illumination changes are more severe.
    - outdoor conditions, many movements, Presence of objects such as buildings, etc.





Drone panoramic view system
- Approach

- Pre-Selection procedure using chaos-inspired similarity measure (fractal dimension)
  - Clustering correspondence points into NxN grid
  - choose the outstanding correspondence points without degradation that was caused by illumination change

$$CISM(I_A, I_B, s) = \frac{2}{D_B(\bar{\Phi}_{I_A,\Delta I})} - 1$$
$$where\ \ \Delta I = I_A - I_B$$

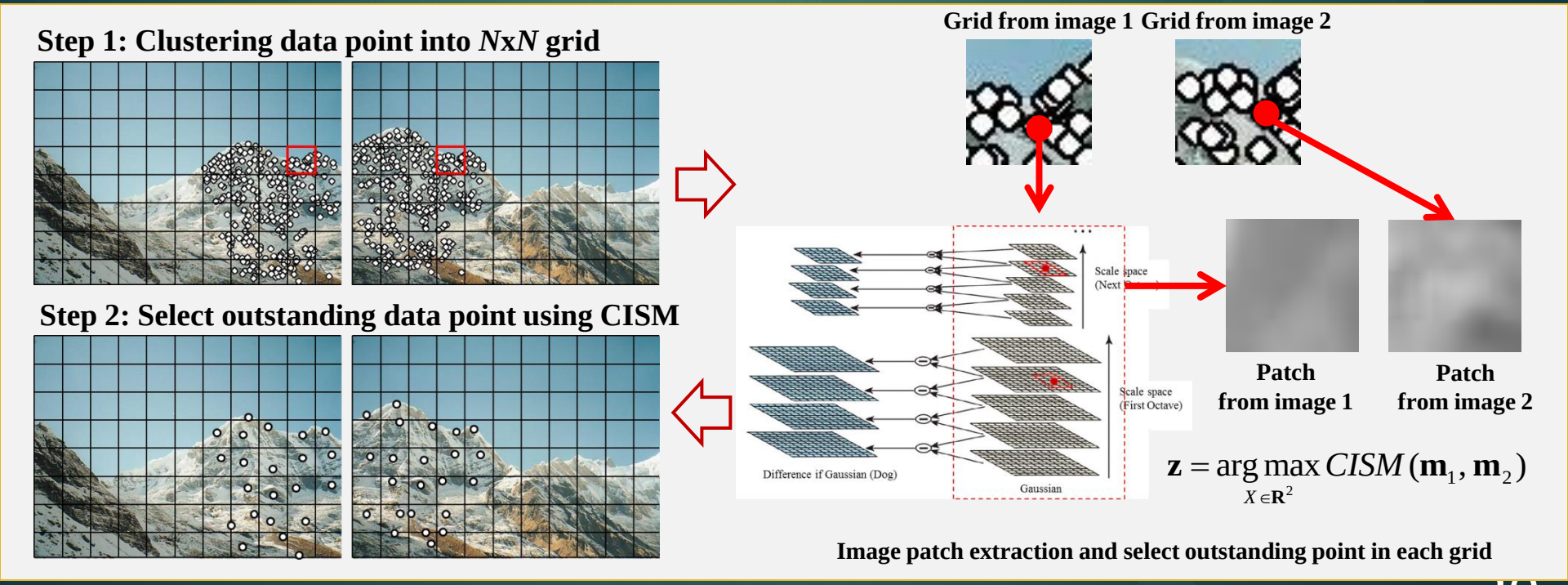
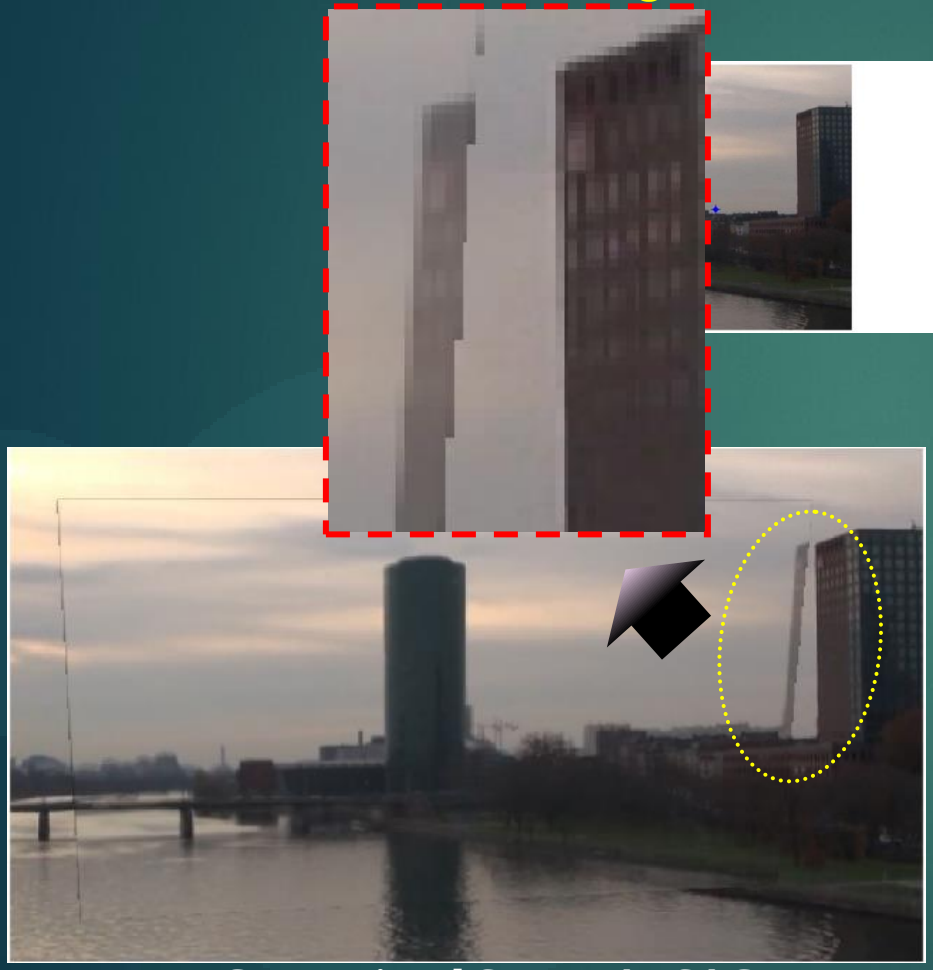
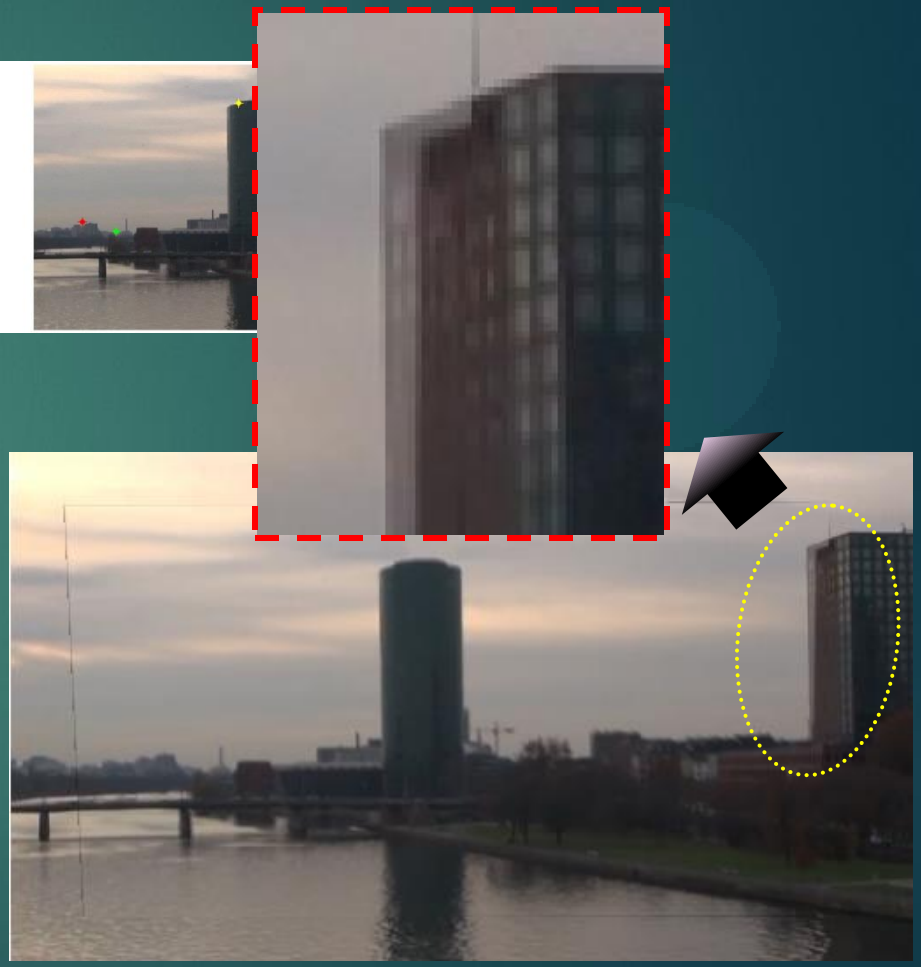


Image Stitching from Drone Platform

- Real-world Drone Image



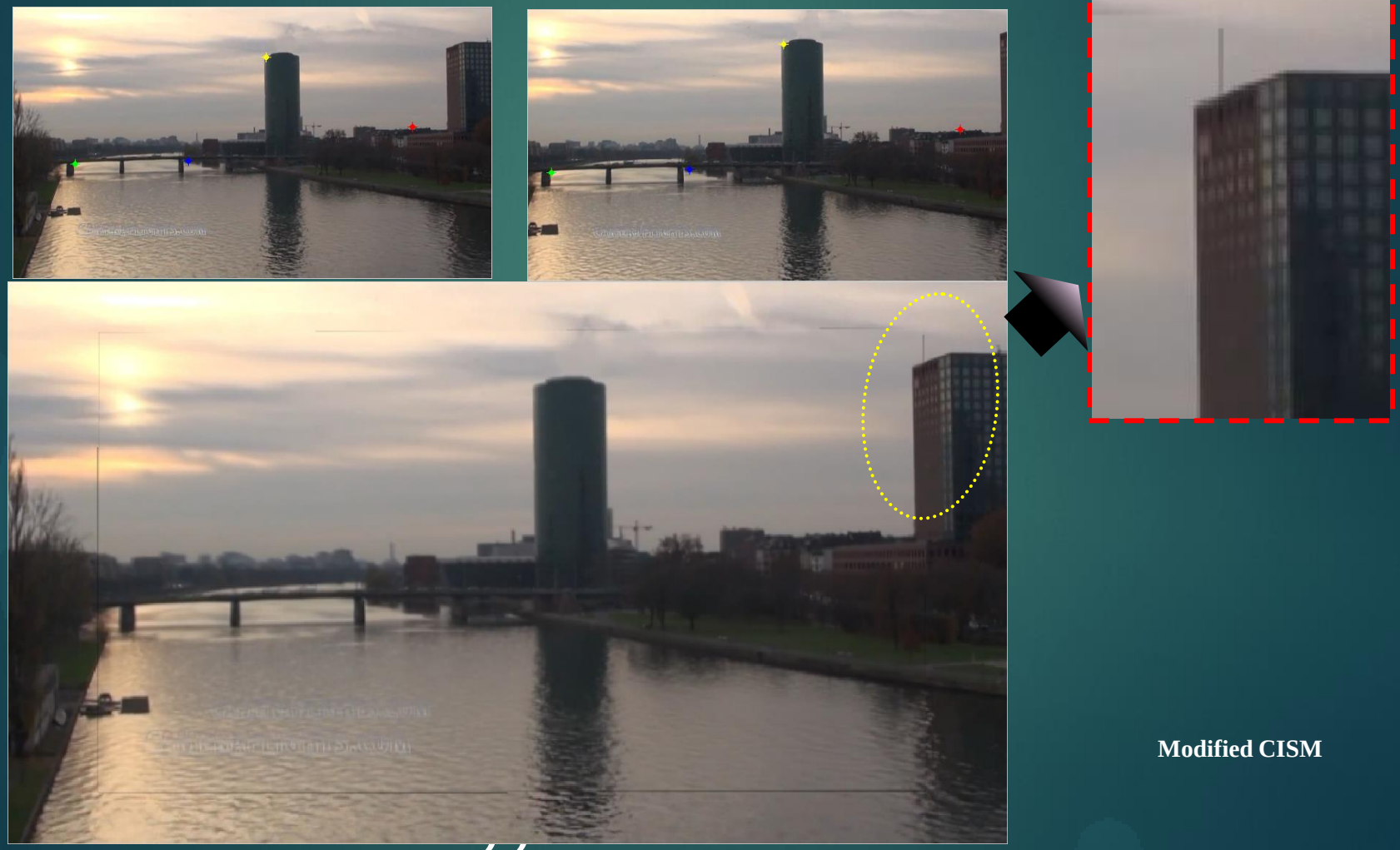
Conventional SIFT-RANSAC



Chaos-RANSAC

Image Stitching from Drone Platform

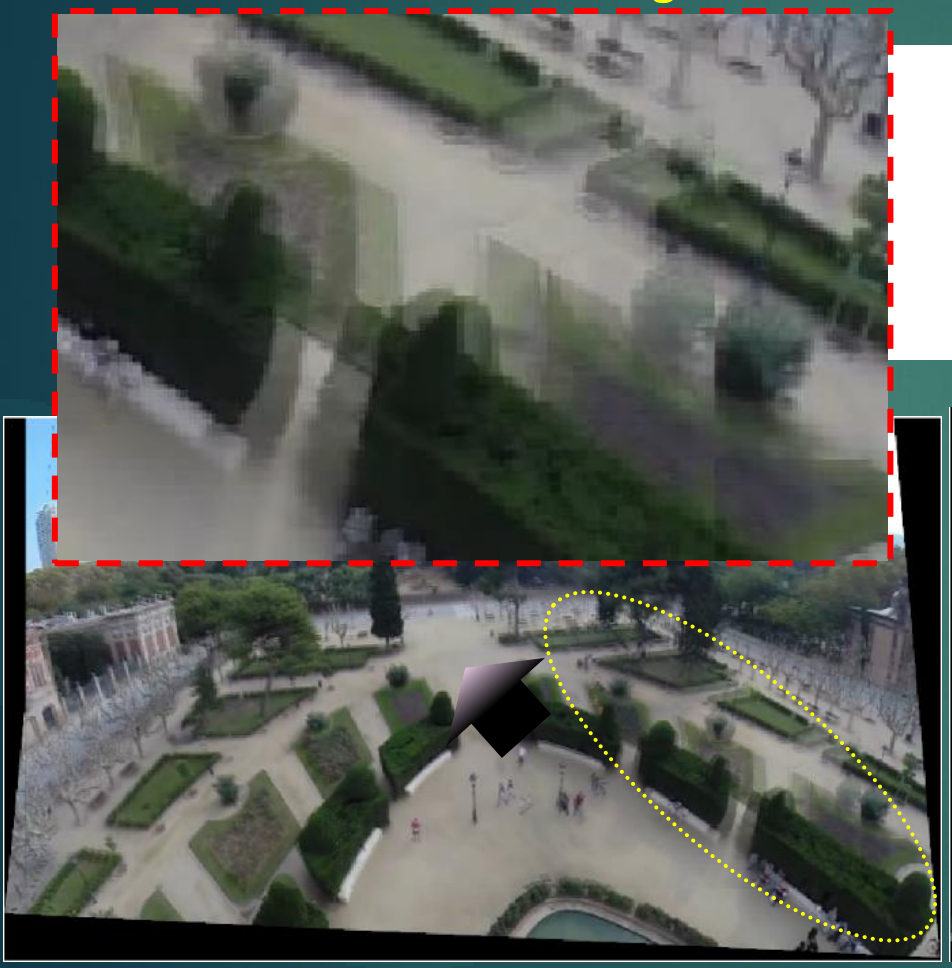
- Real-world Drone Image



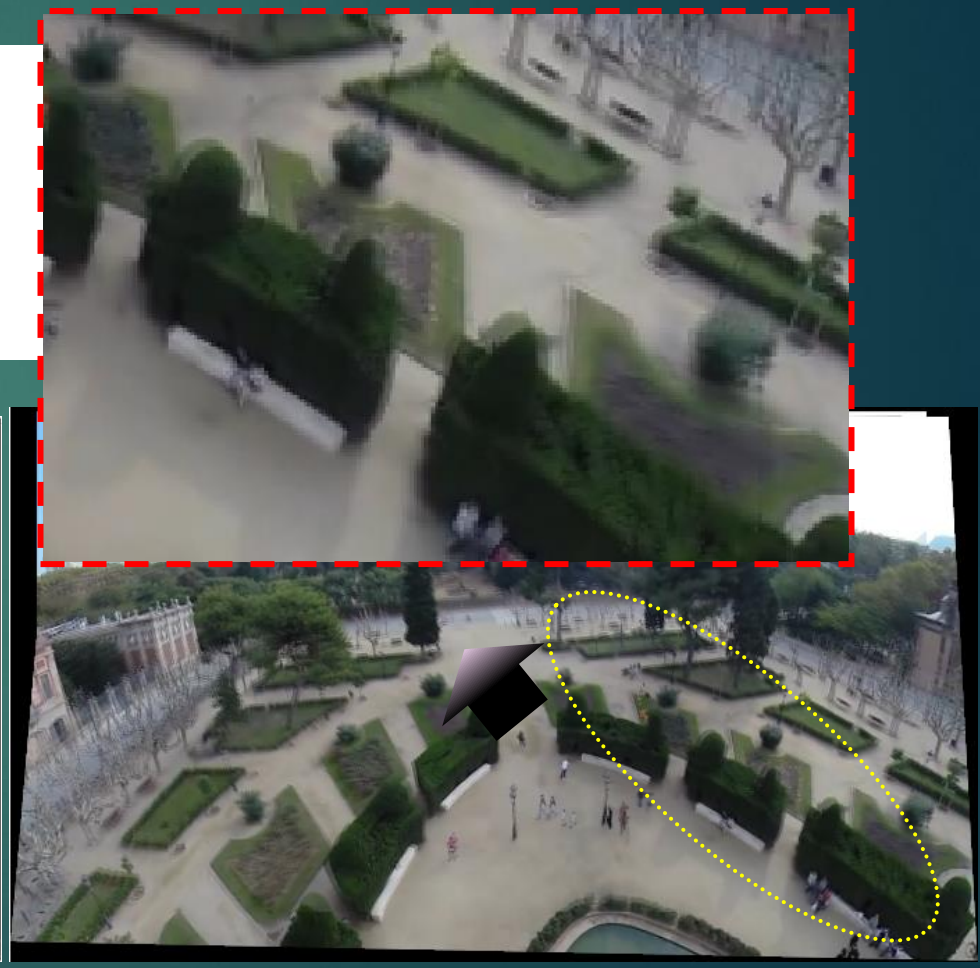
Modified CISM

Image Stitching from Drone Platform

- Real-world Drone Image



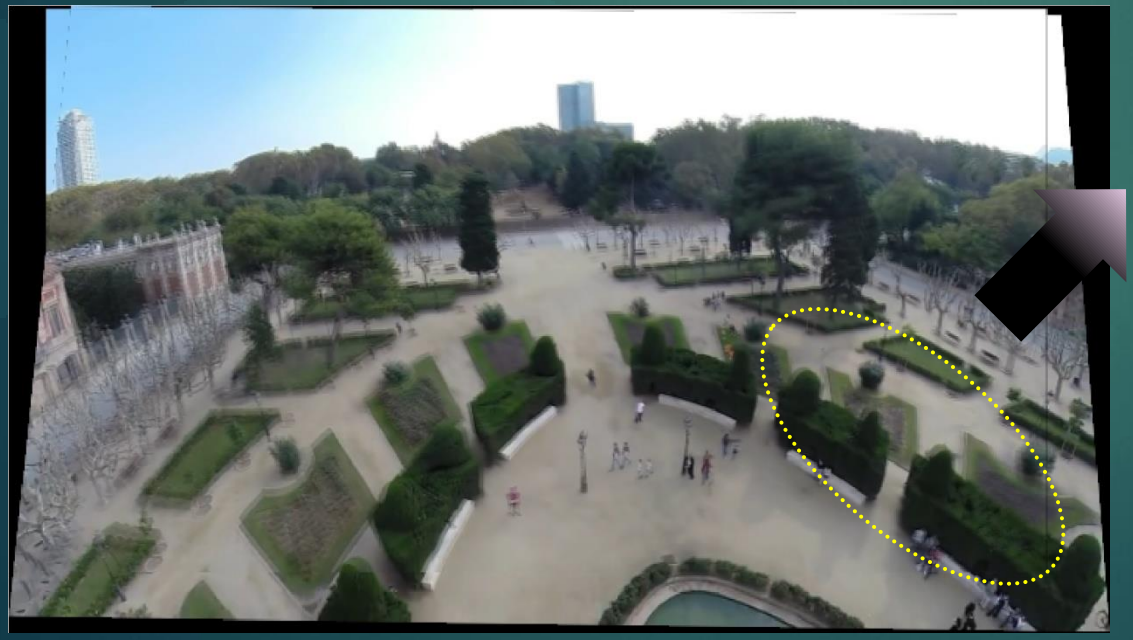
Conventional SIFT-RANSAC



Chaos-RANSAC

Image Stitching from Drone Platform

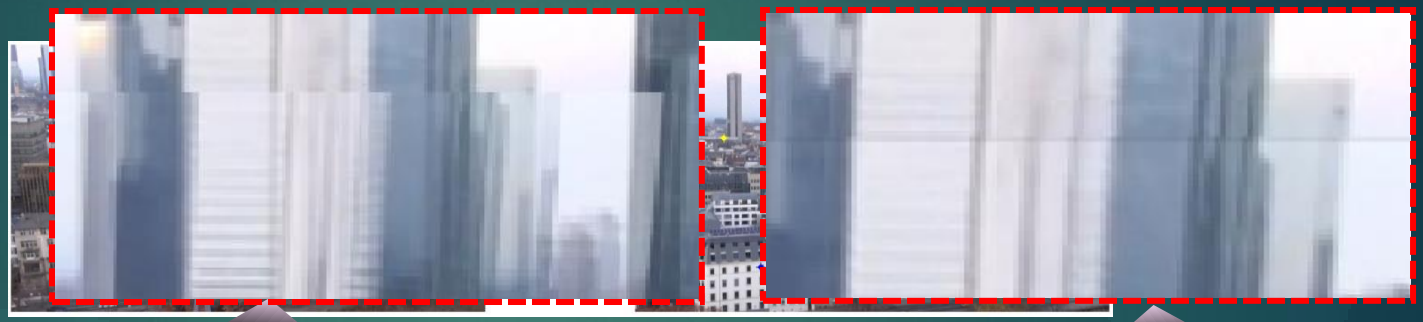
- Real-world Drone Image



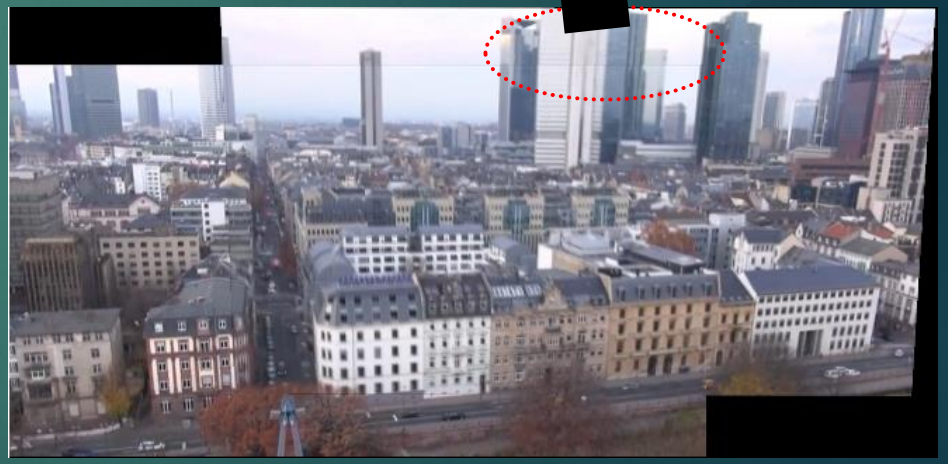
Modified CISM

Image Stitching from Drone Platform

- Real-world Drone Image



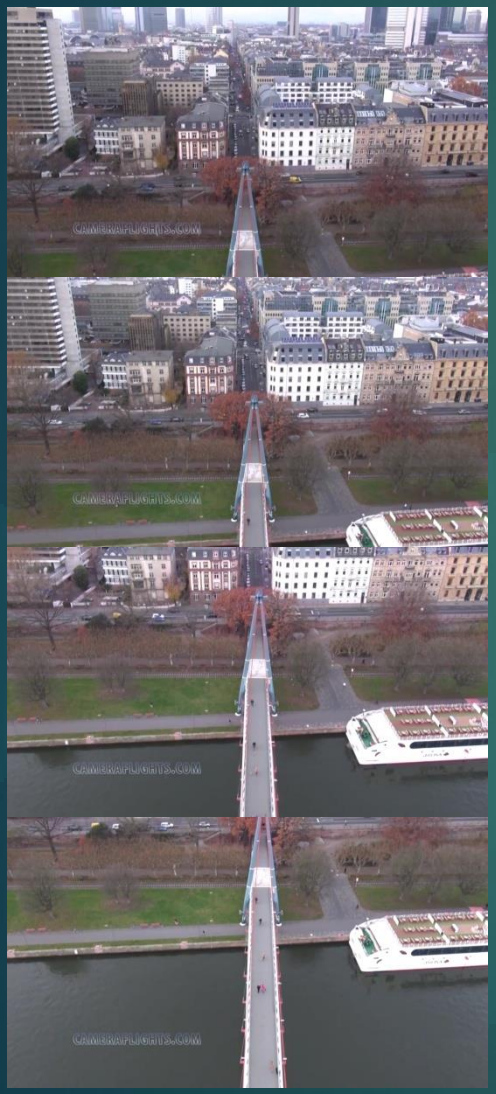
Conventional SIFT-RANSAC



Chaos-RANSAC

Image Stitching from Drone Platform

Image Stitching from Drone Platform



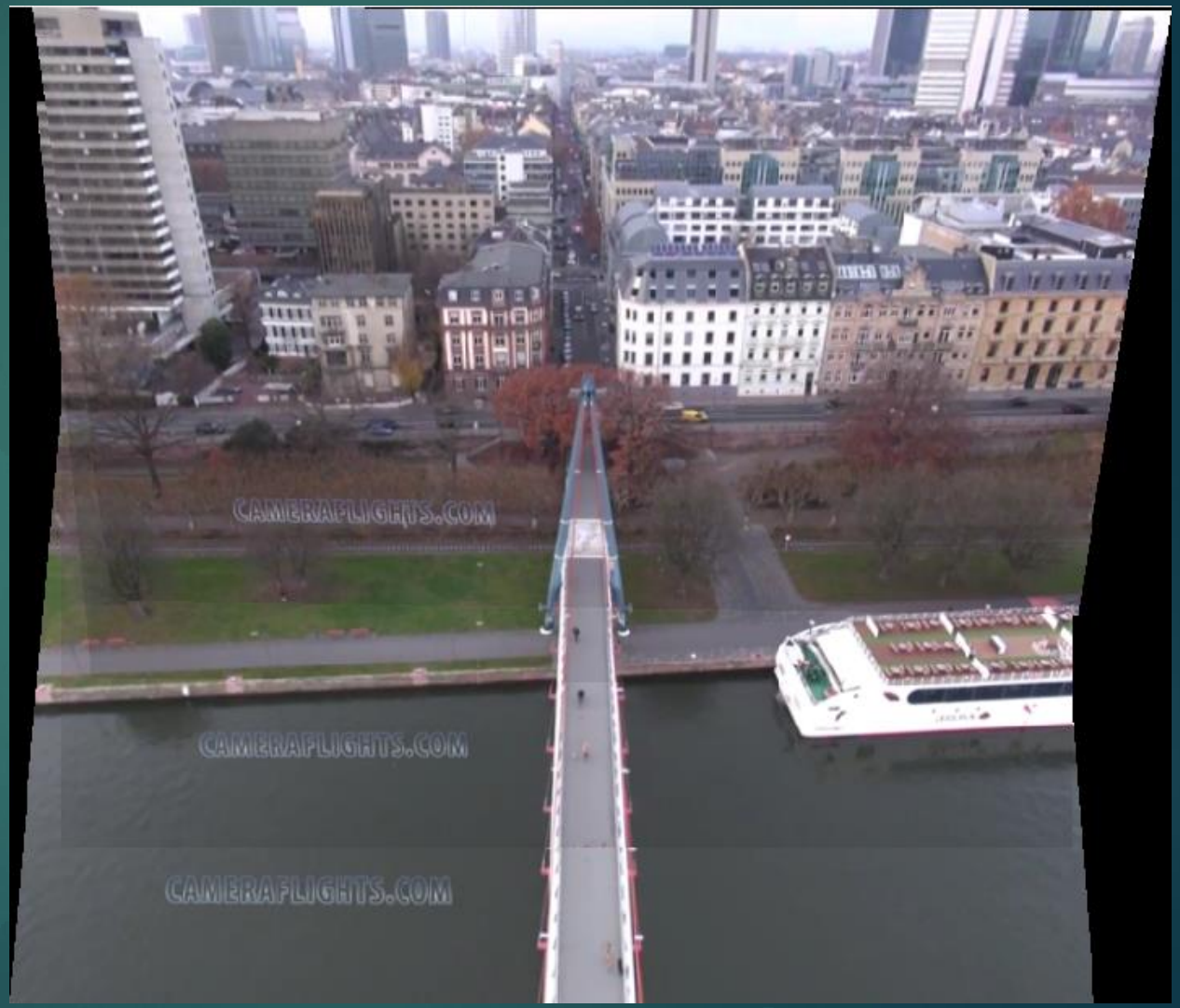


Image Stitching from Drone Platform

Delivering Impact of Weather

AI technology based Video Analytics and Image processing techniques applied to live CCTV feed can deliver the **impact of weather** to citizens and Meteorology Agency **real time** with **timely warnings** and alarms.

Video Analytics, Image Processing

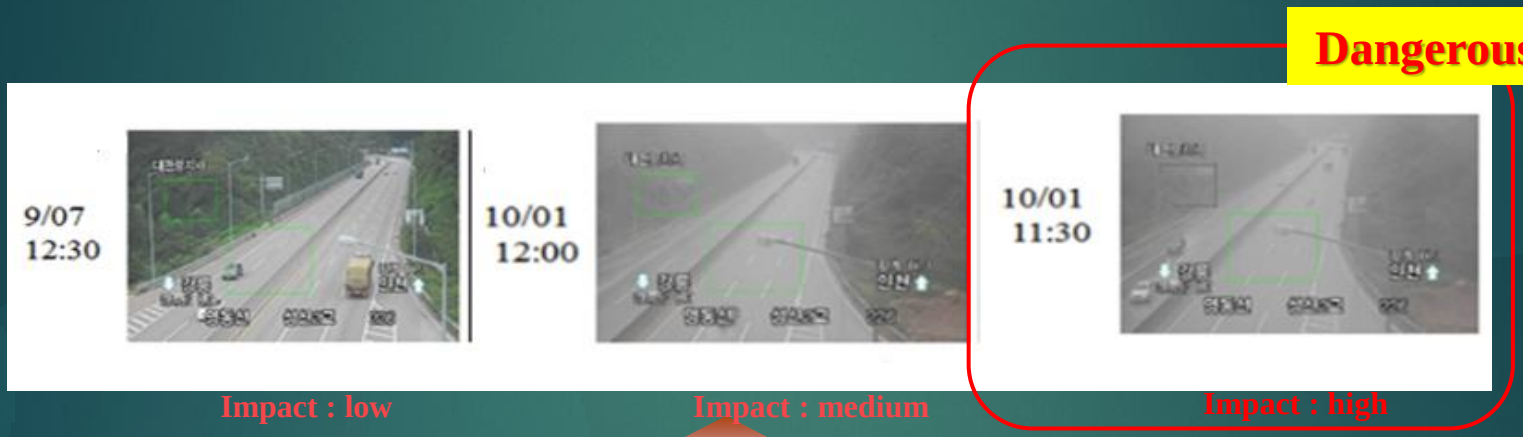
Video Analytic Server

Monitoring Center

Efficiently manage and forecast the impact of weather

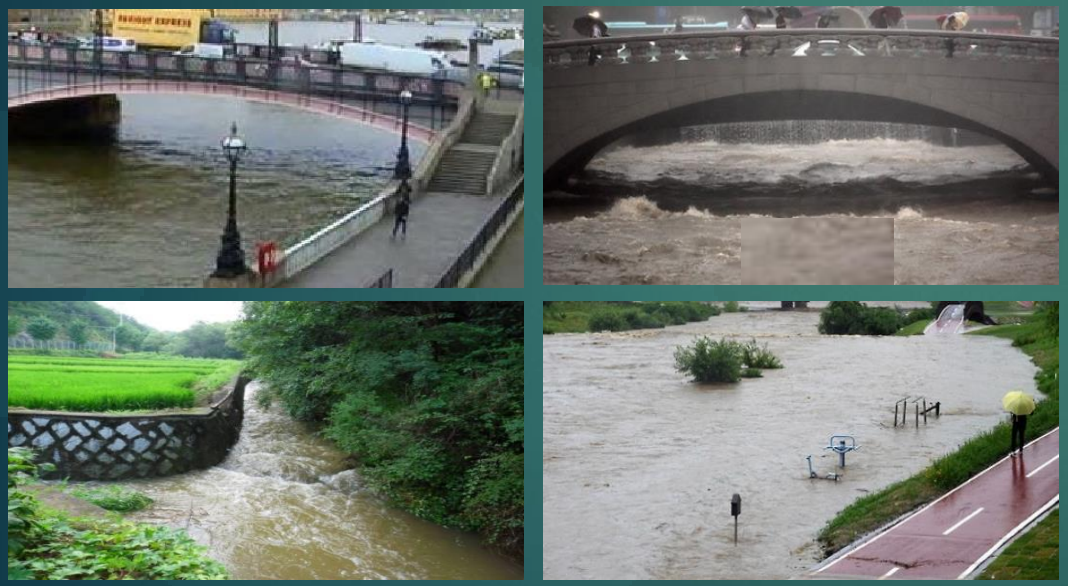
Operational Scenarios under various weather environment

- **Impact of weather on highways using CCTV streams**
  - Managing and forecasting of **fog on highway** can prevent car fatalities (extreme fog, flooding, ice)



Operational Scenarios under various weather environment

- **Impact of weather on river**
  - Managing and forecasting **depth of water in a river** can **prevent possible flooding** and provide timely warning to those being endangered.



< CCTV images from rivers >

Image processing technique

Rainfall raises depth of water in river

Video Analytics

Impact of weather on river

Operational Scenarios under various weather environment

- **Impact of weather on the coast**
  - Managing and forecasting the **wave height** **on the coast** can **prevent damages** from tsunamis.

< CCTV images from the coasts >

Image processing technique

Wave height

Big data Analytic

Impact of weather on the coast

Operational Scenarios under various weather environment

- **Impact of weather on the street**
  - Managing and forecasting the **snow or rain on the street**
  - Impact of weather can also be estimated by **analyzing the activities of pedestrian or cars**

Amount of rain → **Impact of rain**

Number of umbrellas → **Impact of rain**

Amount of snow → **Impact of snow**

Moving pattern of pedestrians → **Impact of condition of road**

Approaches

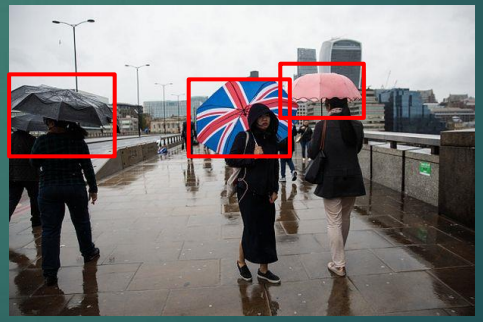
Fixed CCTV based:

people activity monitoring on streets

- **Weather information from fixed CCTV camera on streets**
  - **Clothing** classification: **Indication of temperature**
  - **Umbrella** detection: Indication of rain
  - **Mask** detection: Indication of **dust storm**

< Temperature > ↓

< Temperature > ↑



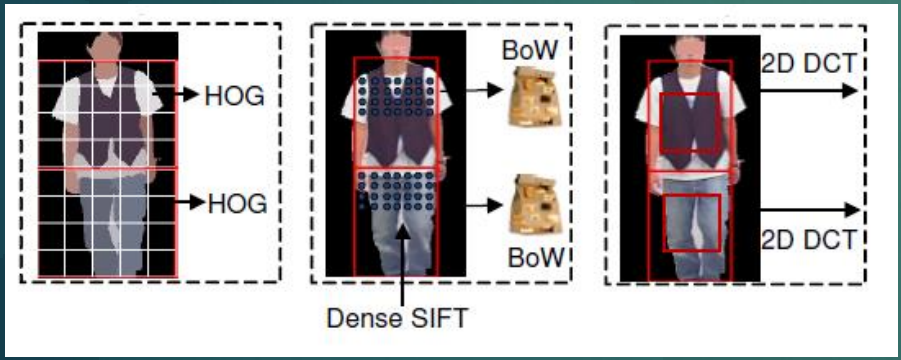
< Rain>

< Dust storm>

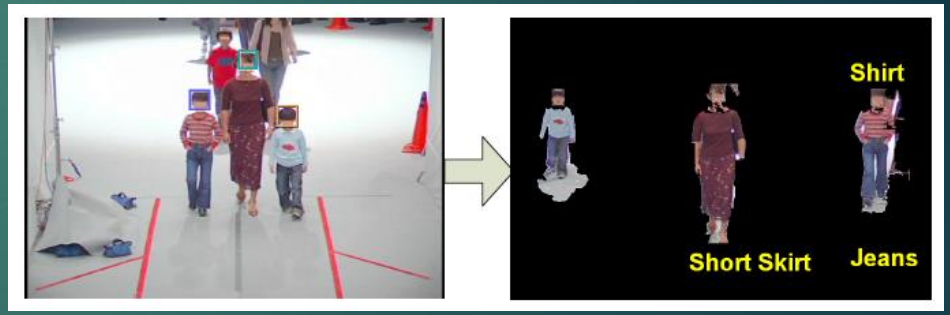
Fixed CCTV based:
people activity monitoring on streets

- **Clothing classification & recognition**
  - Color segmentation method with 3 types of texture features (HOG, BOW, DCT) [1]

HOG: Histogram of Oriented Gradient
BOW: Bag of-Words
DCT : Discrete Cosine Transform



< 3 types of texture features >

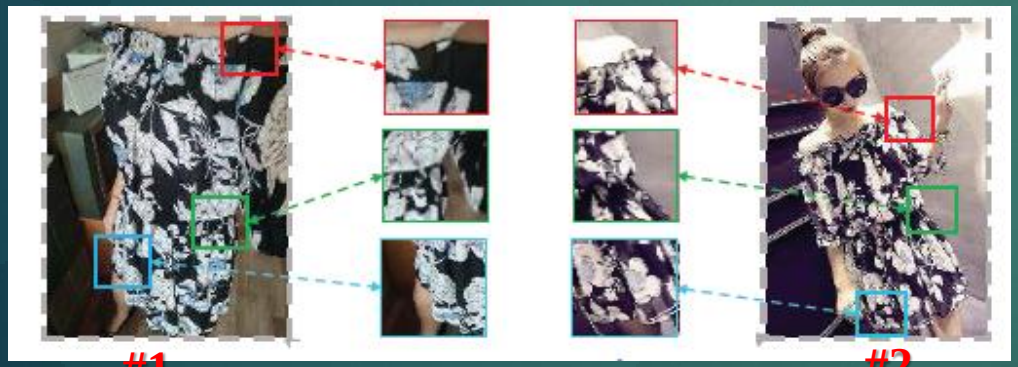


< Recognition example >

[1] YANG, Ming; YU, Kai. "Real-time clothing recognition in surveillance videos". ICIP, pp. 2937-2940, 2011

Fixed CCTV based:
people activity monitoring on streets

- **Clothing classification** (Cont'd)
  - Recently, **deep learning model** (named FashionNet) based on CNN has been developed [2]
  - It learns clothing features by jointly predicting clothing attributes and land-marks



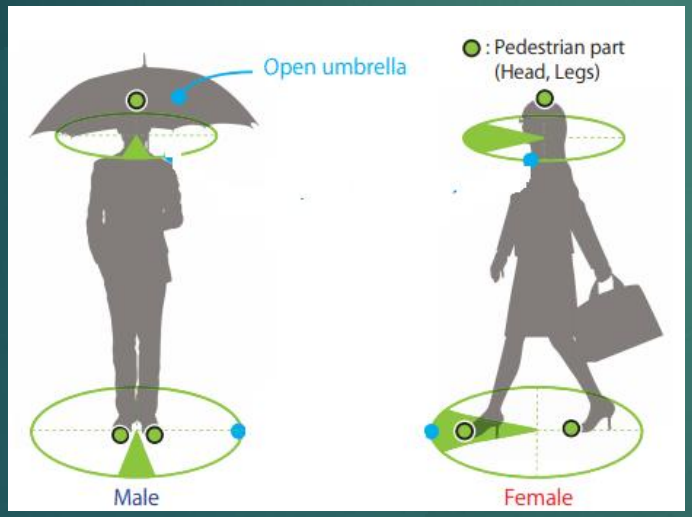
< Additional landmark locations>

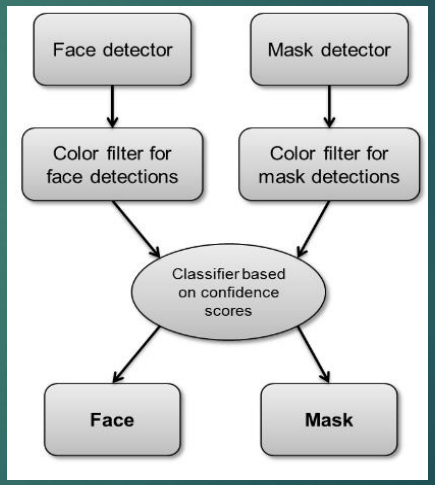
< Partition of clothing feature space>

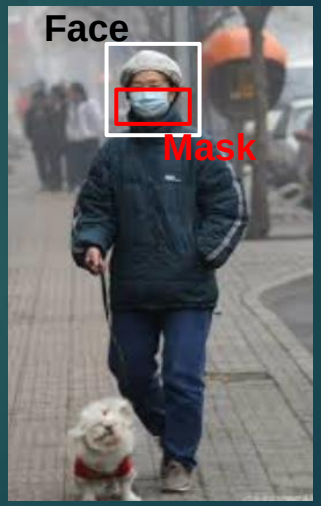
[2] LIU, Ziwei, et al. "Deepfashion: Powering robust clothes recognition and retrieval with rich annotations". CVPR, pp. 1096-1104, 2016

Fixed CCTV based:
people activity monitoring on streets

- **Umbrella / Mask detection**
  - Recognizing the attribute (open umbrella) was proposed in [3]
  - Confidence measures was proposed for the mask and maskless face detections [4]







<Recognizing attributes for open umbrella >

< Mask detection example>

[3] FUKUI, Hiroshi, et al. "Robust pedestrian attribute recognition for an unbalanced dataset using mini-batch training with rarity rate". In: Intelligent Vehicles Symposium pp. 322-327, 2016

[4] N.Rodriguez, al Mask and maskless face classification system to detect breach protocols in the operating room. International Conference on Distributed Smart Cameras. ACM, pp. 207-208, 2015

Fixed CCTV based:

augmented weather surveillance

- **Extreme Weather Detection and Alarm**

<Normal Weather>

<Extreme Weather>

Dense fog, Severe rain/snow

Moving CCTV based:
on vehicle weather impact collection

- **Extreme Weather detection and alarm for safety**

<Extreme Weather Management on the road>

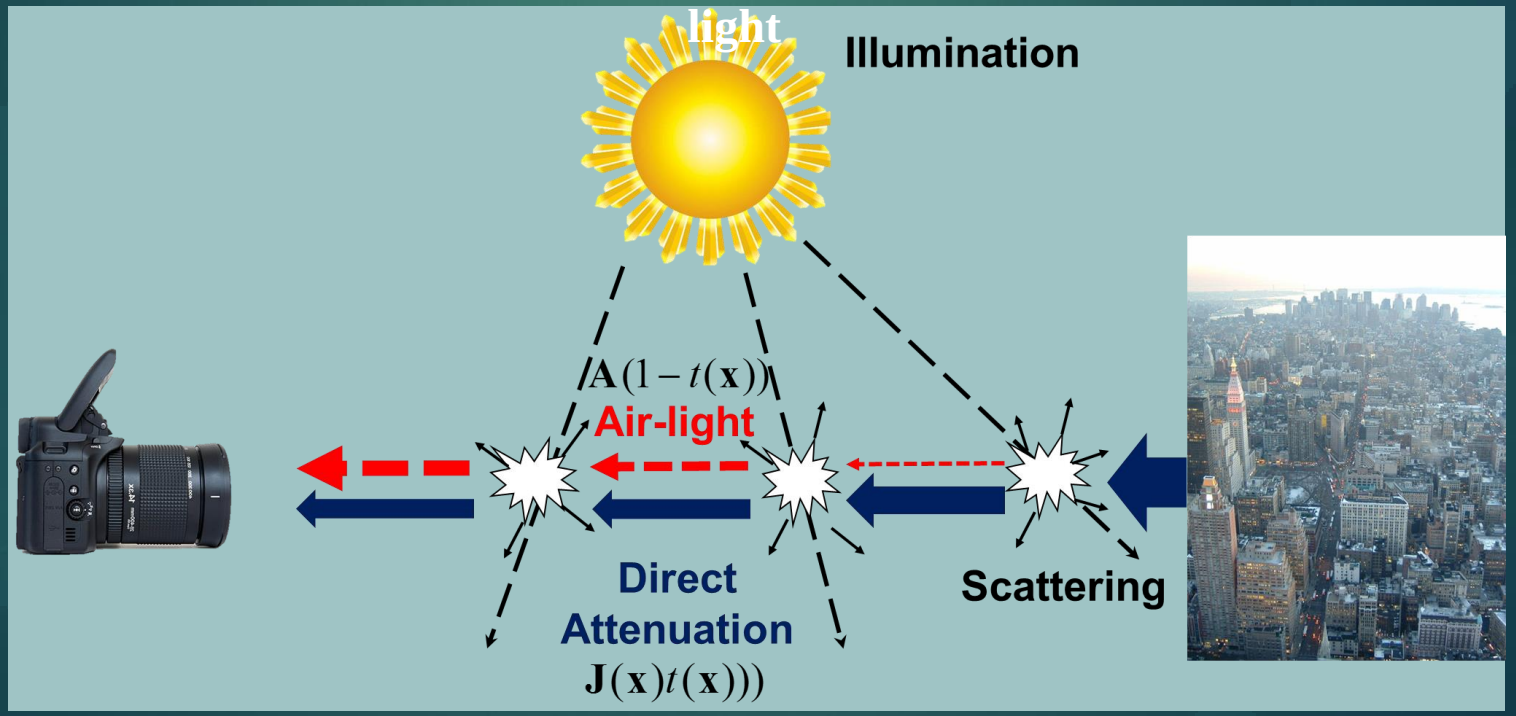
Proposed Algorithms

Measuring Fog Density

- Fog Imaging Model
  - Optical Imaging and Image Representation

$$I(x) = J(x)t(x) + A(1 - t(x))$$

I(x): Fog image, J(x): Fog-free image, t(x): Transmission, A: Atmospheric



Measuring Fog Density

- **Grouping fog density**
  - **Group 0**: Clear, **Group 1**: Very Thin, **Group 2**: Thin
  - **Group 3**: Normal, **Group 4**: Thick, **Group 5**: Very Thick

<Outdoor Image Database>

Measuring Fog Density

- **Feature for Measuring Fog Density**
  - Fog Feature derived from fog imaging model
  - Feature 1

$$d^{\mathbf{I}}(x) = \left(\min_{n \in \{r,g,b\}} \mathbf{J}^n(x) \right) t(x) + A_0 (1 - t(x)),$$

  - Feature 2

$$b^{\mathbf{I}}(x) = \left(\max_{n \in \{r,g,b\}} \mathbf{I}^n(x) \right) t(x) + A_0 (1 - t(x)),$$

  - Feature 3

$$c^{\mathbf{I}}(x) = \left(\max_{n \in \{r,g,b\}} \mathbf{J}^n(x) - \min_{n \in \{r,g,b\}} \mathbf{J}^n(x) \right) t(x),$$

$$= c^{\mathbf{J}}(x) t(x),$$

Measuring Fog Density

- **Global Atmospheric Light**
  - Global atmospheric light constraint

$$\bar{b} \le A_0 \le \max b^{\mathrm{I}}(x),$$

  - Initial global atmospheric light

$$A_0 = \lambda \max b^{\mathrm{I}}(x) + (1-\lambda)\bar{b},$$
$$0 \le \lambda \le 1$$

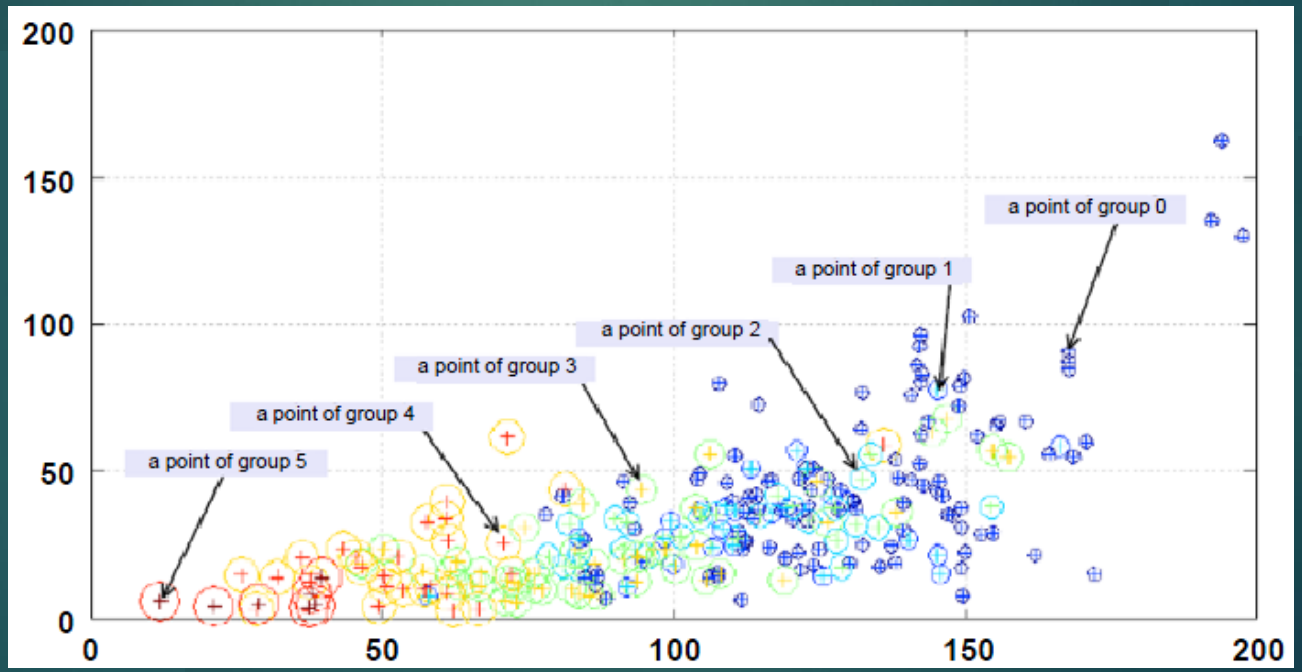
(λ : control para. atmospheric light)

Measuring Fog Density

- **Distribution according to the fog density**
  - Feature distribution map

$$A_0 - \bar{d}(x) = (A_0 - \bar{d}(x))t(x),$$

$$\bar{c} = \frac{1}{M}\sum d^{\mathbf{I}}(x) - \frac{1}{M}\sum b^{\mathbf{I}}(x),$$



Measuring Fog Density

- **Fog density estimation function**

$$\omega = \exp\left\{-\frac{1}{2}(5.1\alpha + 2.9\beta) + 0.2461\sigma\right\},$$

$$\alpha = \frac{A_0 - \bar{d}}{A_0}, \beta = \frac{\bar{c}}{A_0}$$

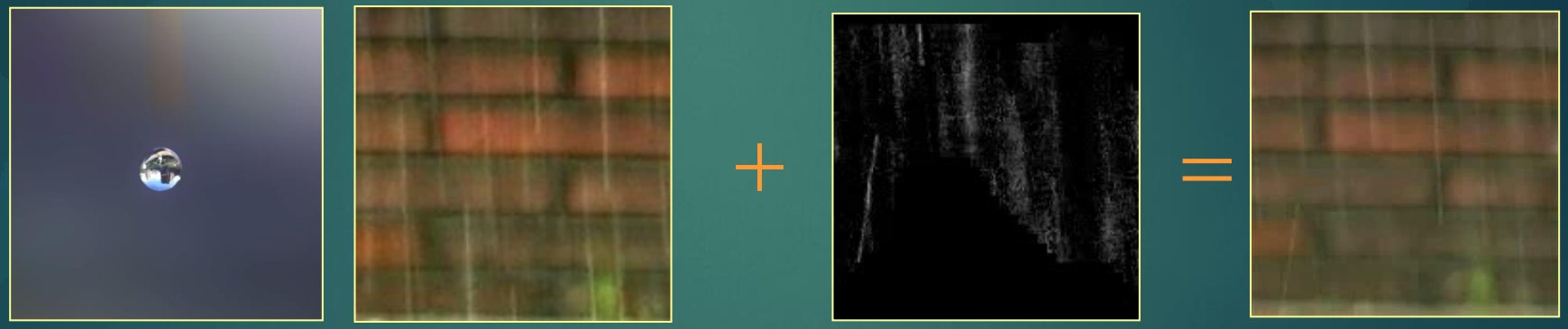
Fog density	ω	Class	
		Ratio of fog	Fog density
0	0.1	0	Clear
1	0.3	30~50%	Low
2	0.5	50~60%	Thin
3	0.7	60%~70%	Normal
4	0.8	70%~80%	High
5	0.9	80%~100%	Denger

Measuring Fog Density

- **Demo**
  - Fog density estimation

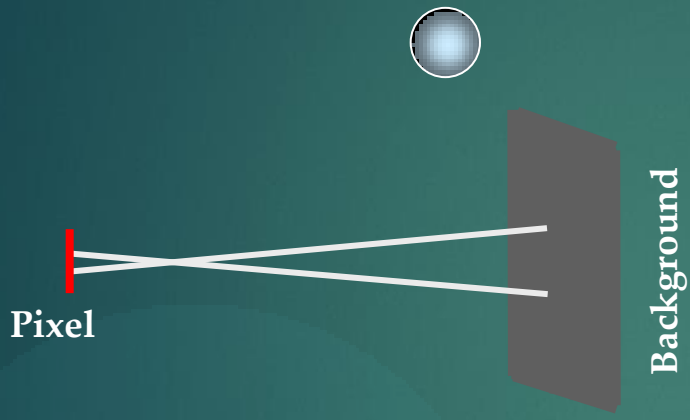
Measuring Snow/Rain

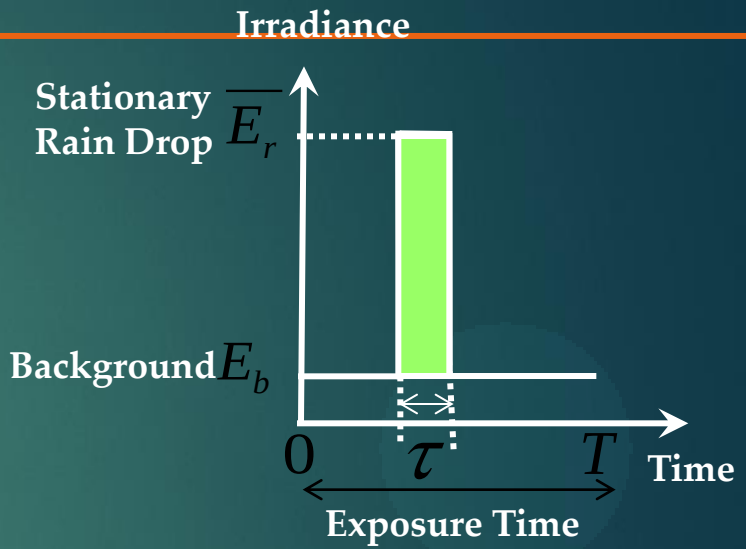
- **Characteristics of Snow/Rain**
  - Rain/snow degrade the visibility of the scene and disrupt precise analysis
  - Rain/snow have the dynamics and occur sporadically



Measuring Snow/Rain

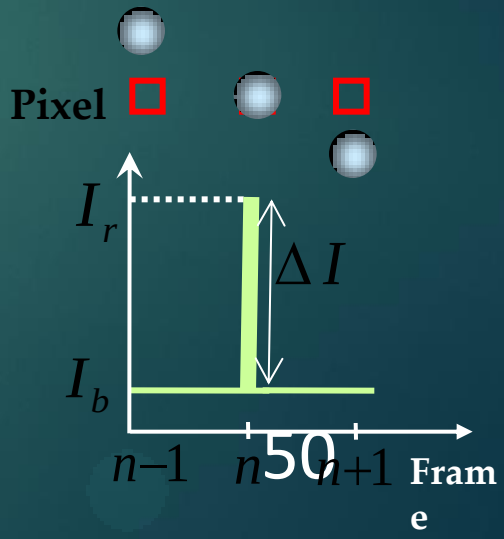
- Rain/Snow Imaging Model





Time a drop stays over a pixel : $\tau < 1\,ms << T$

Change in Intensity : $\Delta I = \tau(\overline{E_r} - E_b) > 0$

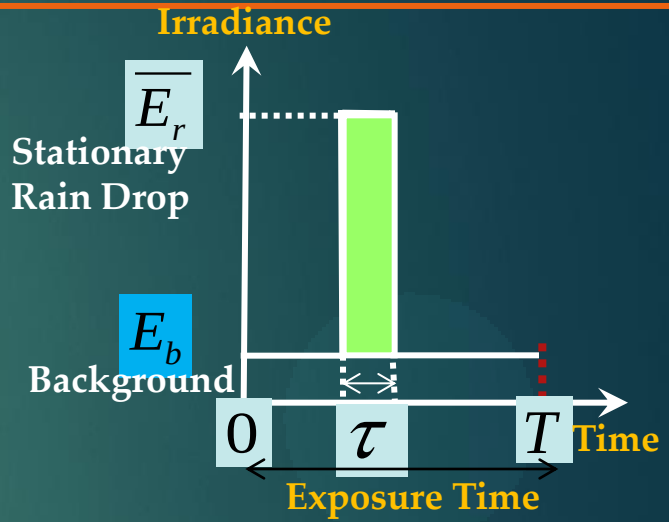


Rain produces a "Delta Signal" in Time

Measuring Snow/Rain

Intensity at a Pixel on a Rain Streak :

$$I_r = E_b(T - \tau) + \tau \overline{E_r}$$



Intensity of the Background :

$$I_b = E_b T$$

Change in Intensity due to a Rain Streak :

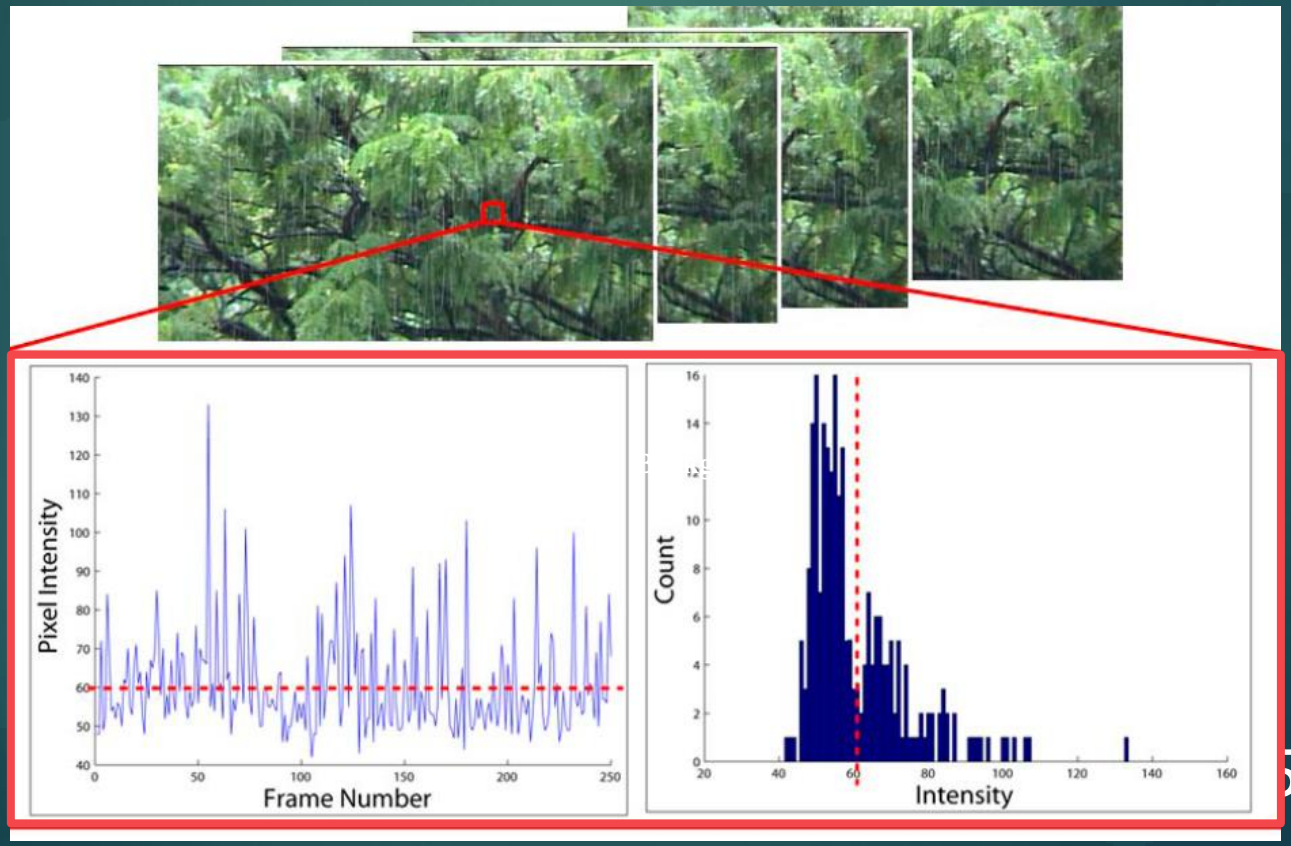
$$\Delta I = -\frac{\tau}{T} I_b + \tau \overline{E_r}$$

Constants Constants

Change in Intensity is Linearly related to Background

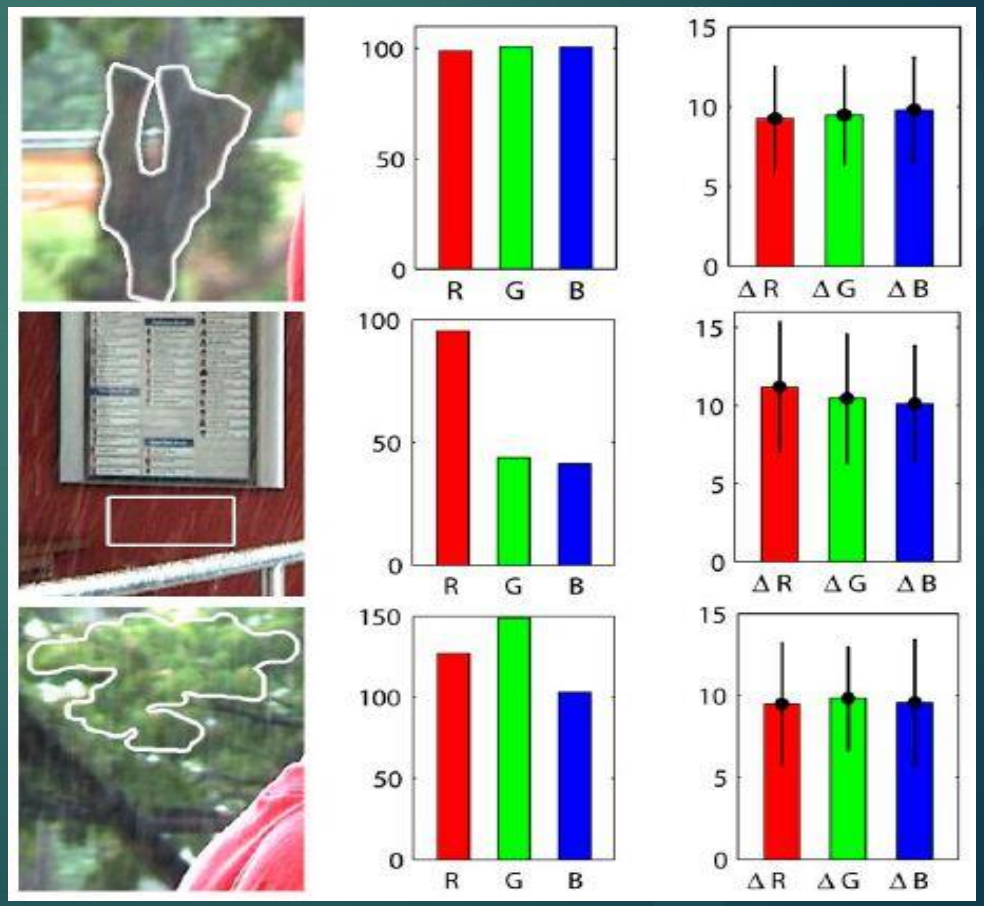
Measuring Snow/Rain

- **Rain/Snow Detection**
  - The pixels including rain drop have the 2-peaks in the temporal histogram
  - 2-peaks are computed by K-means clustering
  - Classify the pixels based on 2-NN method using 2-peaks



Measuring Snow/Rain

- **Rain/Snow Detection**
  - Initially detected pixels are further classified using chromatic constraint which the differential of each R,G,B channels is almost same regardless of background color



Measuring Snow/Rain

Rain detection

- Initially detected pixels are further classified using chromatic constraint which the difference of each R,G,B channels is almost same regardless of background color

Rainy image

Initial detected rain

Detected rain with chromatic constraint

54

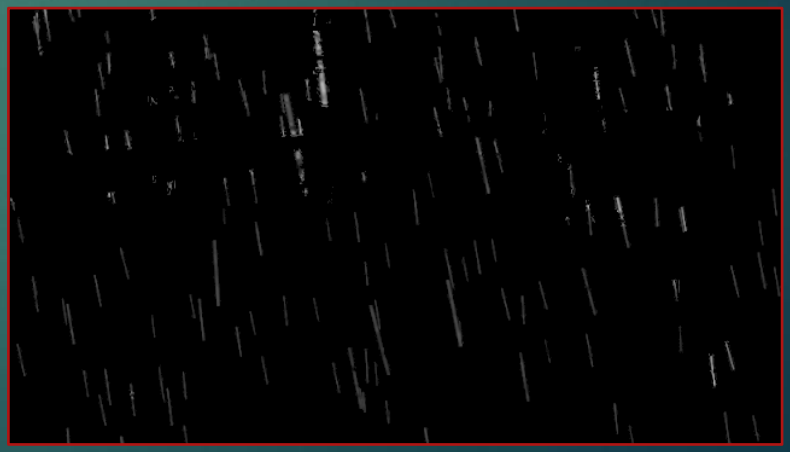
Measuring Snow/Rain

- Experimental result

Rain image

Detected rain

Rain image

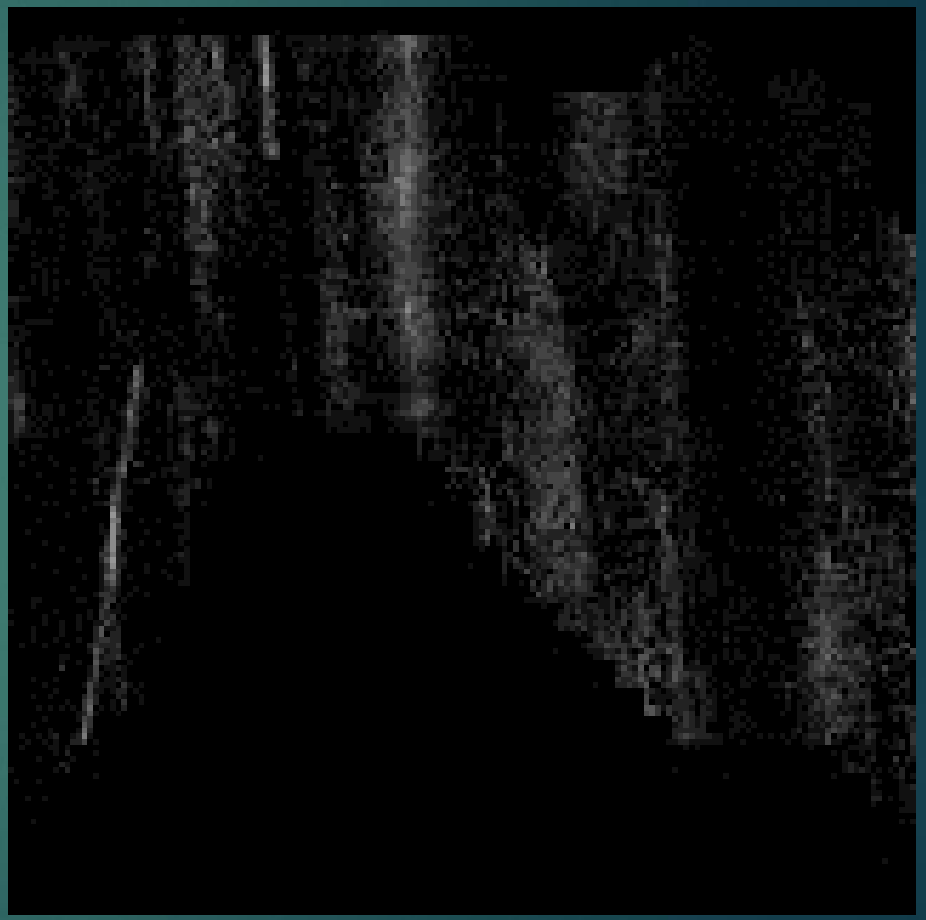


Detected rain

Measuring Snow/Rain

- Demo

Rain Video

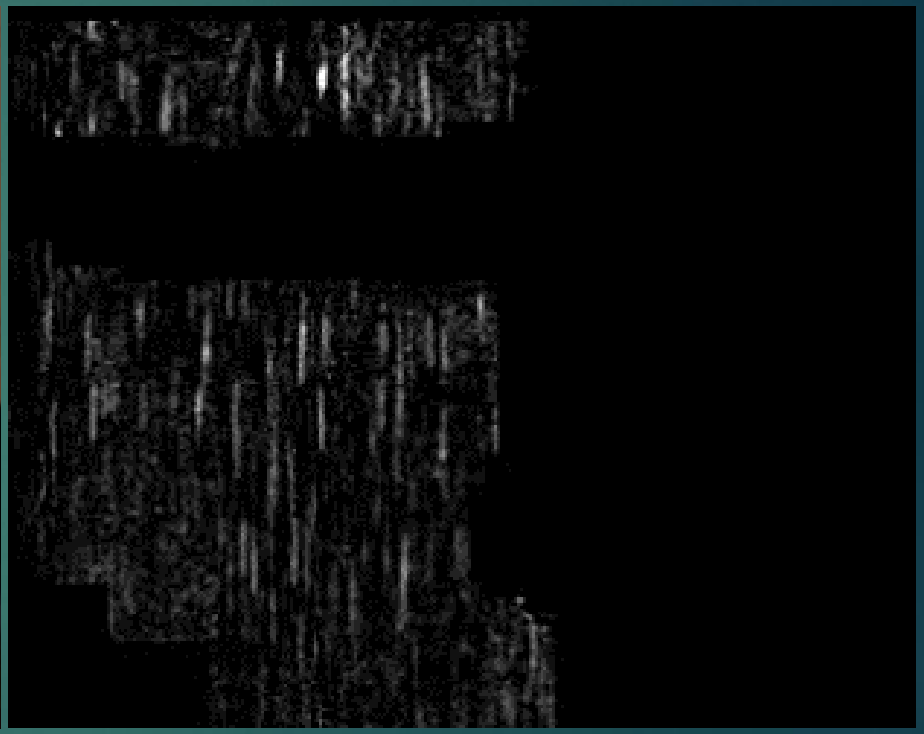


Detected rain

Measuring Snow/Rain

- Demo

Rain Video



Detected rain

Measuring Snow/Rain

- Demo

Snow Video

Detected Snow

58

Human activity recognition

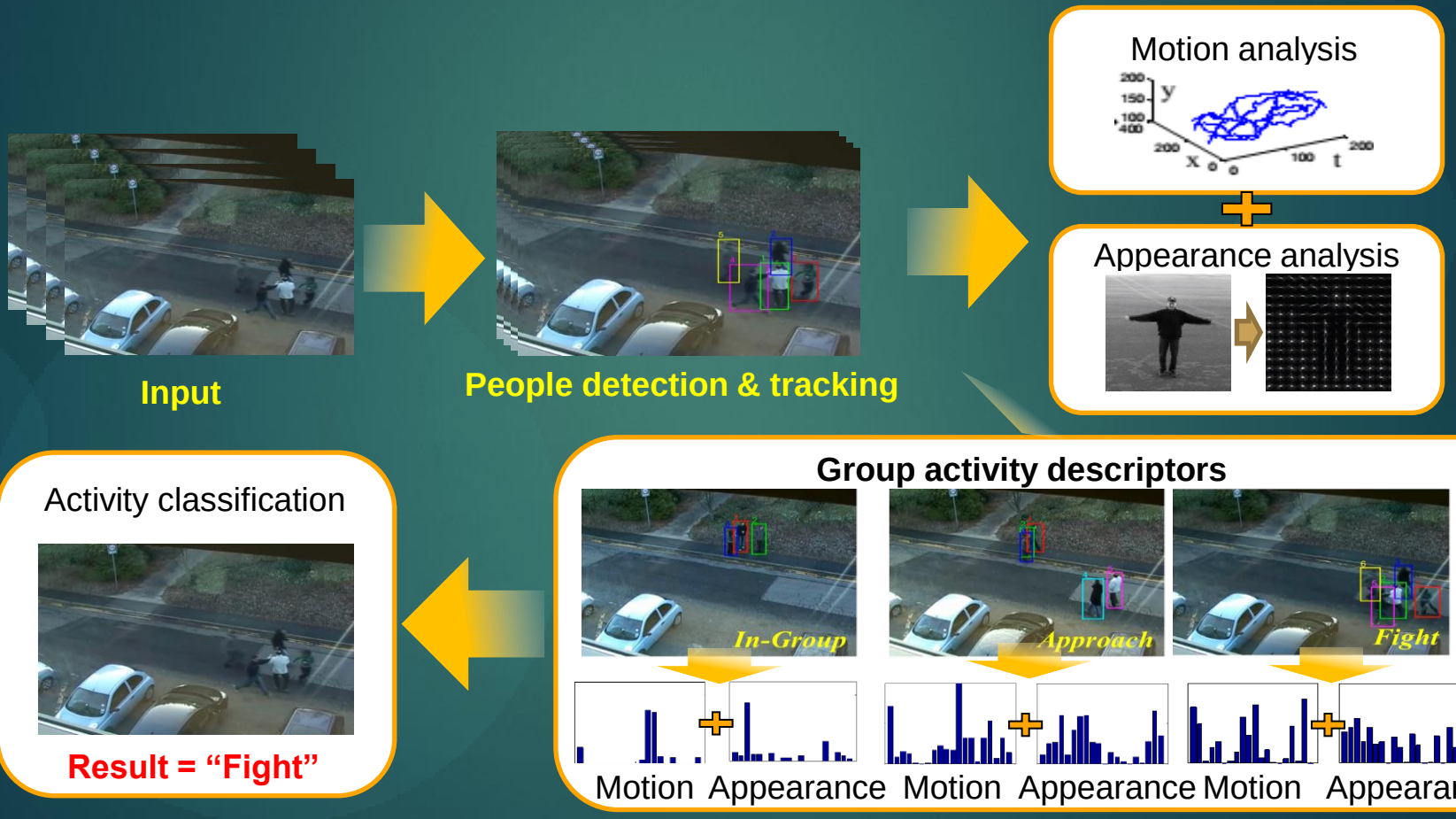
- Human activity recognition is an important and challenging task for video analysis and understanding.

<Example of human activities >

- The analysis of group activity is of great practical importance for many applications.
  - smart video surveillance, semantic video indexing

Human activity recognition

- **Proposed approach**



Human activity recognition

Motion-based Individual activity descriptor

- **Average and variance of velocity**

$$\bar{v}_i = \frac{1}{t_e - 1} \sum_{\tau=1}^{t_e-1} \|v_i(\tau)\|$$

$$\rho_i = \frac{1}{t_e - 1} \sum_{\tau=1}^{t_e-1} \left| \|v_i(\tau)\| - \bar{v}_i \right|^2$$

  - $v_i(\tau) = s_i(\tau + 1) - s_i(\tau)$: the velocity of *i*-th trajectory at frame τ
  - It represents the intensity and variation of the movement

- **Feature vector of motion-based Individual activity descriptor**

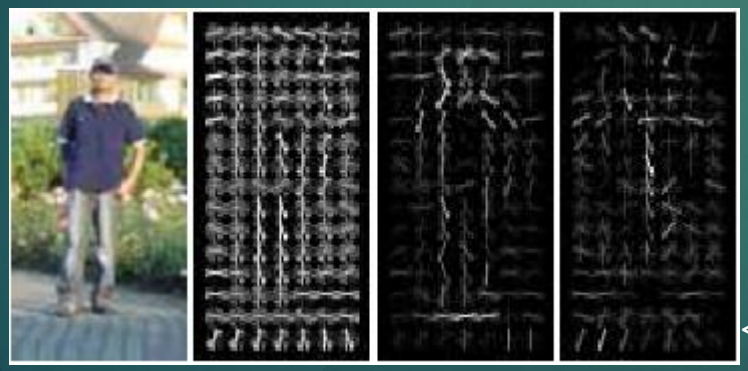
$$[\zeta_i, \bar{v}_i, \rho_i]$$

  - ζ_i : the location change of *i*-th person
  - $\bar{v}_i$: the average velocity of *i*-th person
  - ρ_i : the variance of velocity for *i*-th person

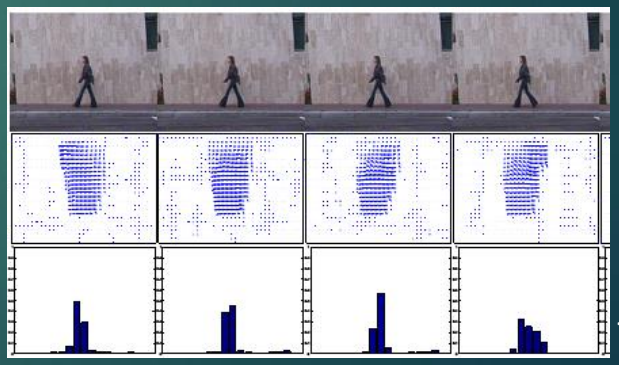
61

Human activity recognition

- **Appearance-based Individual activity descriptor**
  - Appearance information such as shape and local motion is also an important clue for recognizing group activity such as *fight*.
  - Histograms of Oriented Gradients (HOG) & Histogram of Oriented Optical Flow (HOOF) has been employed
    - The sequences of HOG and HOOF features within the bounding box of each trajectory are extracted.



<HOG>



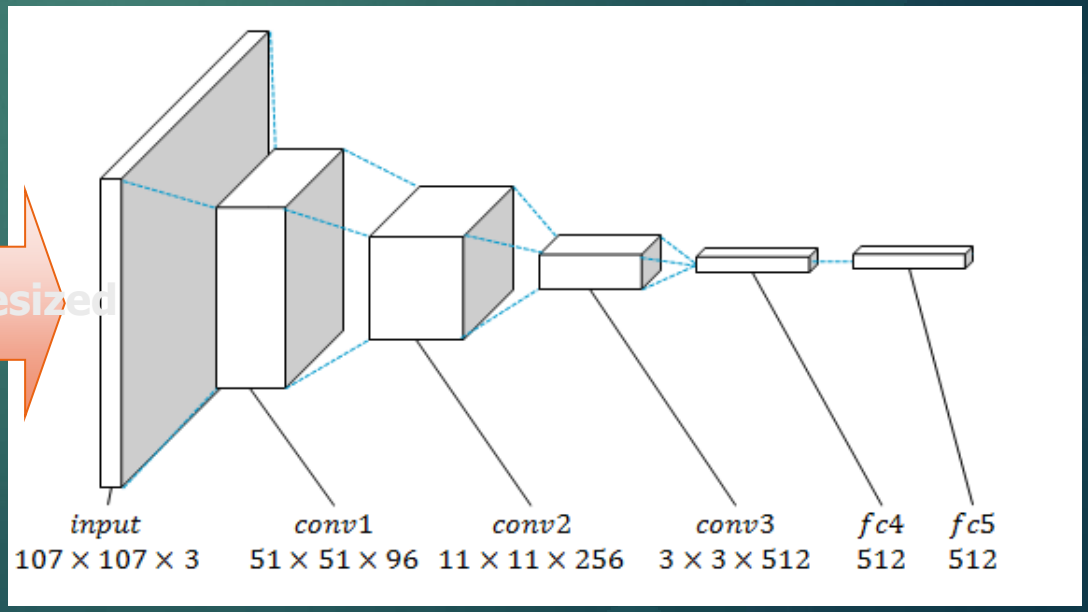
<HOOF>

62

Human activity recognition

- **Appearance-based Individual activity descriptor**
  - Recently, deep learning based descriptors for effectively representing target appearance has been presented
  - Appearance feature vector is compute by output of the proposed multi target network based on CNN

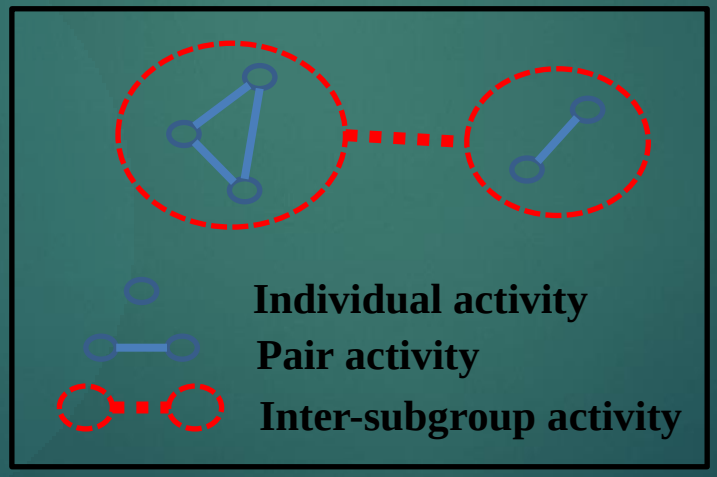
< Detection Results >



< Proposed multi-target CNN >

Human activity recognition

- **Group activity descriptor**
  - Group activity is regarded as a combination of subgroups and characterized by three types of activity descriptors
    - Descriptor type : “individual”, “pair”, “inter-subgroup”
    - These are complementary for analyzing group activity



<Three types of group activities>

Human activity recognition

- **Group activity descriptor**
  - Consider the interactions between different sub-groups
  - It is useful in complex situations with multiple group behavior such as *approach* or *split.*
  - Individuals are clustered into non-overlapping sub-groups using the mean-shift clustering algorithm based on their position.
    - No need to determine the number of groups in prior

Human activity recognition

- **Activity classification with group activity descriptors**
  - All activity description features are extracted and accumulated from the training video samples.
  - For every type of features, the visual words are generated using K-means clustering.
  - A multi-class SVM classifier are trained using a MKL method [5] which combines various types of these features and feed the mixture of the kernels to SVM.

[5] P. Gehler, S. Nowozin, "On feature combination for multiclass object classification", ICCV, pp. 221–228, 2009

Human activity recognition

- Demo

67

Conclusions

- **Novel approaches to measure and forecast the impact of weather based on video analytics and image processing from fixed and moving CCTV are presented**
  - Methods to accurately measure the density of fog, rain and snow, respectively
  - Method to recognize the human activity utilizing deep appearance model and group activity descriptor
- **Proposed approaches can be applied to analyzing the impact of adverse weather**
  - Measuring and Forecasting the impact of weather on highway / street / in river / coast
  - Forecasting the impact of weather based on activity analysis of pedestrian

- Thanks for your attention!

Q & A