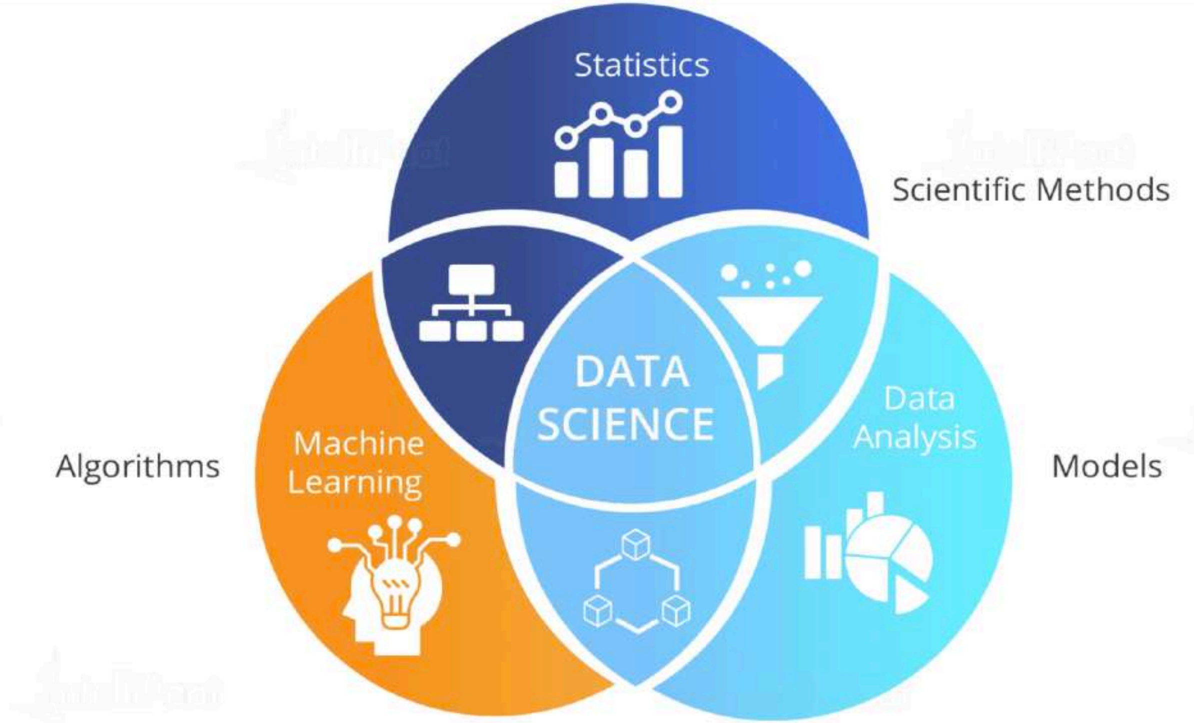
2012

Data Science is the sexiest job in 21st Century

2019

Is Data Science the sexiest job in 21st Century?

Are U Data Scientist?



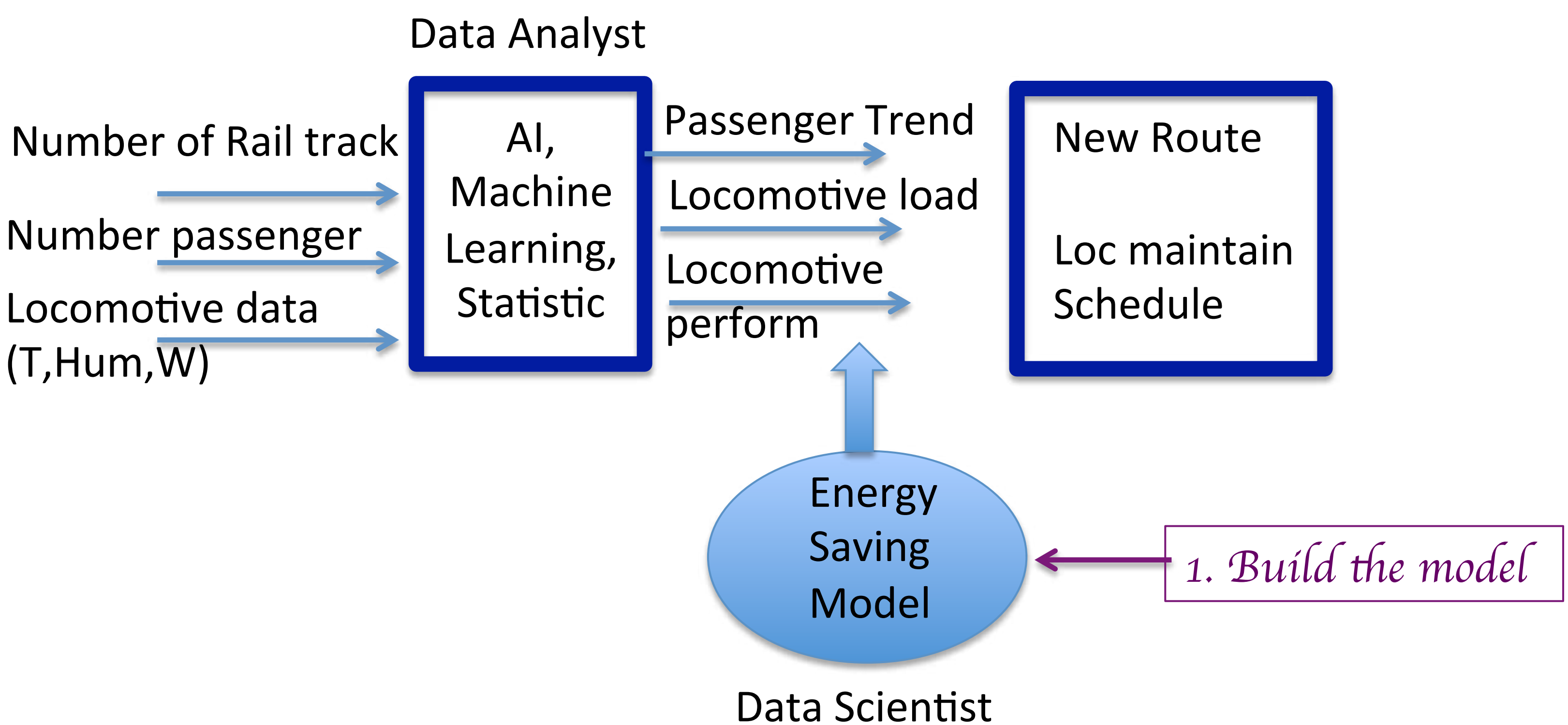
create value from data through analytics

data science is integral to the business product, vision and success

Are U Data Scientist?

GE Transportation Use Case

Business Purpose : Help rail companies manage locomotives better



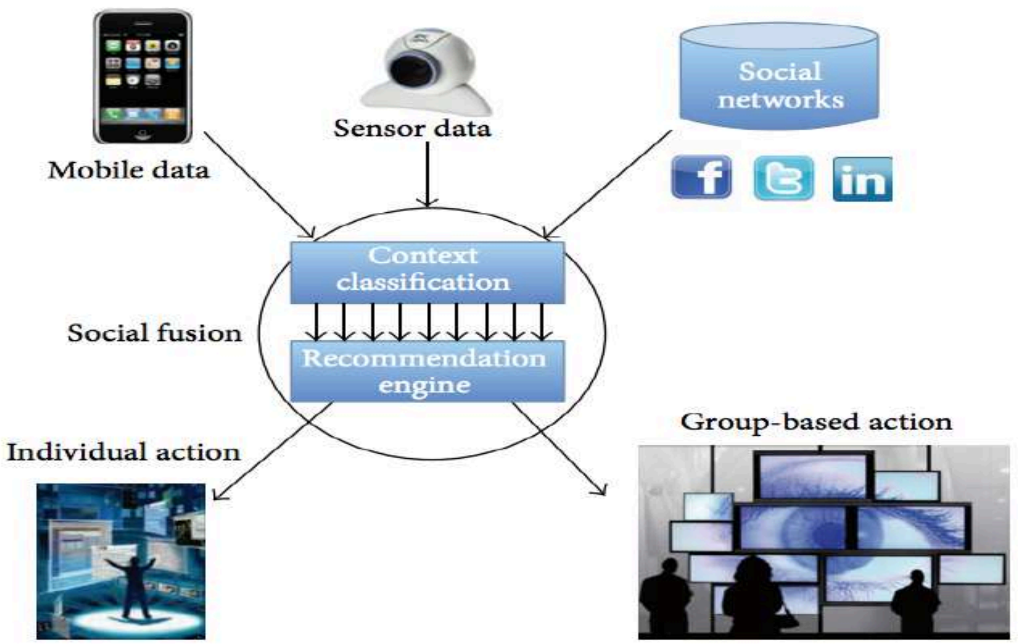
The Sexiest…

1. Deal With Fusion Sensing : Utilize many sensor to draw the whole plant

Remote sensors, satellites, and unmanned aerial vehicles (commonly referred to as "drones") can monitor plant health, soil conditions, temperature, nitrogen utilization, and much more around the clock.

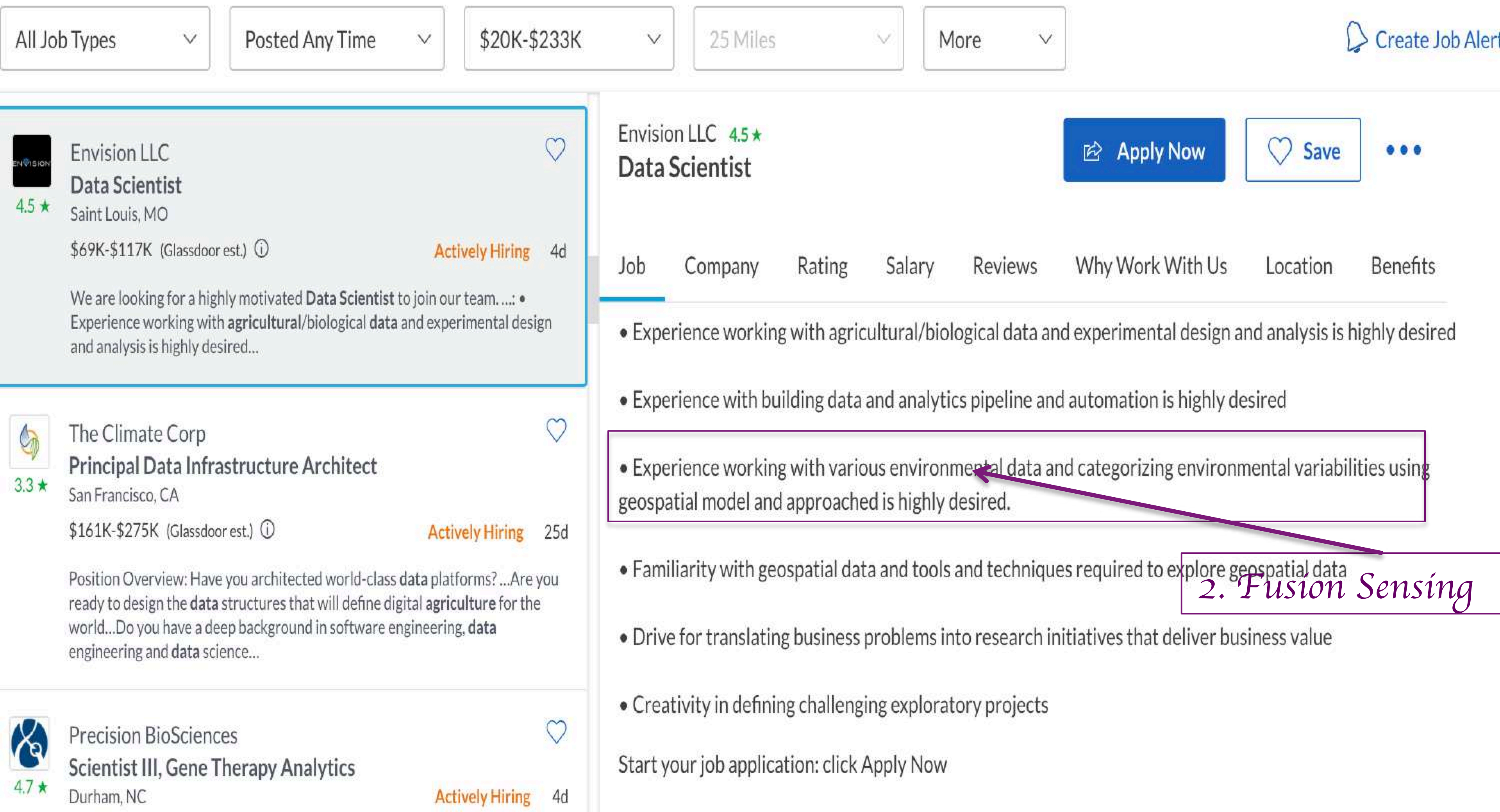
The Sexiest…

1. Deal with Fusion Sensing : Utilize many sensor to build the Knowledge



Platform to Build the Knowledge Base by Combining Sensor Data and Context Data
Sungho Shin, Jungho Dongmin Seo,Sung-Pil Choi, Seungwoo Lee,
Hanmin Jung, and Mun Yong

For the Sexiest…



The Sexiest…

2. Understanding Outlier

Observation which deviates so much from other observations as to arouse suspicion it was generated by a different mechanism….. Hawkin (1980)

↓

Novelties

When trying to detect outliers in a dataset it is very important to keep in mind the context and try to answer the question: "Why do I want to detect outliers?" The meaning of your findings will be dictated by the context.

The Sexiest…

2. Understanding Outlier

1. Check Sensor : procedure of measurement, sensor reliability, environment
   - if : all above are ok then remove outlier
   - else : follow procedure bellows

2. Try in different condition
   - if : there are no more outlier then finish
   - else : follow procedure bellows

3. Check the model : fixed restriction/assumtion
   - if : no more outlier then finish
   - else : throw away all data but outlier. Change the model

The Sexiest…

3. Build the distinguish model :
employ unusual method from different field

Viral Marketing Model : Random Brownian (Biology Molecular)

THE MODEL

Movement

$$m\frac{dv}{dt} = F_{ran}(t) - \gamma v(t)$$

$$v(t_{n+1}) = v(t_n) + (-\gamma v_n\, dt + dW)/m$$

$$dW = F_{random-force}$$

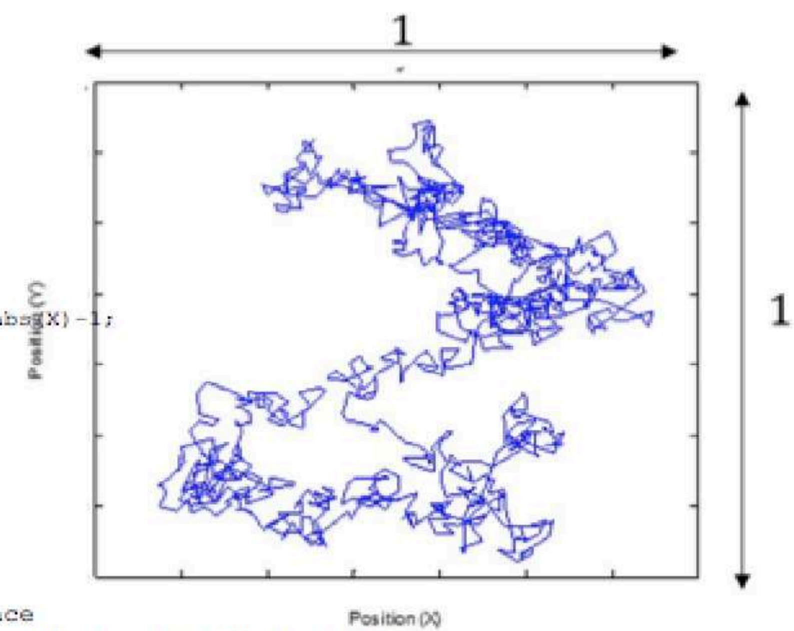
$$= Gaussian\ distribution$$

$$= \sqrt{2m\gamma kT}\, dt\ x\ Gaussian\,[0,1]$$

$$Gaussian[0,1] = \sqrt{-2\ln(r_1)} x\cos(2\pi r_2)$$

Rule ➡ **Reflection of the trajectory**

```
When trajectory (X) <0 or X >1
1. Find ΔX; ΔX = X − mx1 while (X)<0||(X)>Bx   X=abs(X)-1;
num=num+1;
X=abs(X);   End
2. Reflect ΔX; for m even and odd
numo=mod(num,2);   if numo==0
X=X;
else   X=Bx-X;
end
```



Constant Force

```
if trajectory <boundary of reaction space
velocity=((2*D*dt)^0.5)*dWx + velocity -((X-Inpx)/abs(X-Inpx))*k;
X=X+ velocity *dt;
```

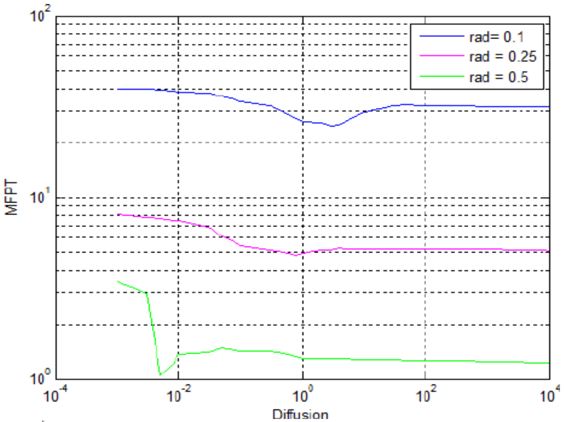
➡ **Spring Force**

```
if trajectory <boundary of reaction space
velocity=((2*D*dt)^0.5)*dWx + velocity -((X-Inpx)/abs(X-Inpx))*(k*(X-0.5));   X=X+
velocity *dt;
```

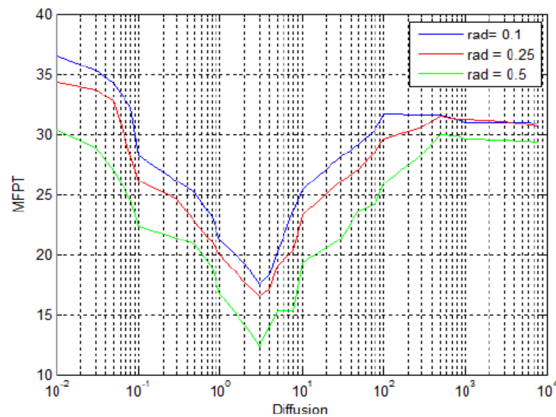
MEAN FIRST REACTION TIME (MFPT)

First reaction time is defined as the time when particle reside in reacting space and react. In viral marketing “reaction” means a viral.

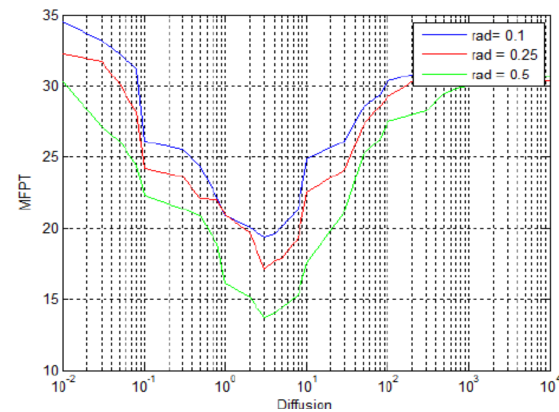
MFPT Without Force.



MFPT With Constant Force.



MFPT With Spring Force.



	Simulation result	Analytical Result
r=0.1	31.6949	31.32
r=0.25	5.0459	5.091
r=0.5	1.2248	1.273

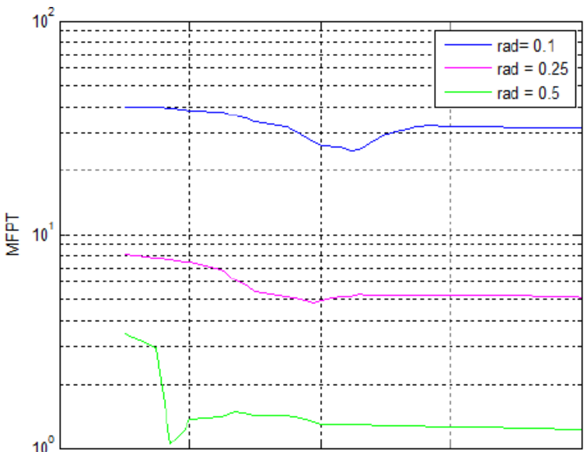
There are anomalous in MFRT in low diffusion. The anomalous, in this study is defined as optimal diffusion coefficient. MFRT is usually increase inversely, but in small range of D, the MFRT is tends to decrease until certain value and increase again. The trend is reinforced by attracting force.

optimal diffusion coefficient represent possibility viral will happened

MEAN FIRST REACTION TIME (MFPT)

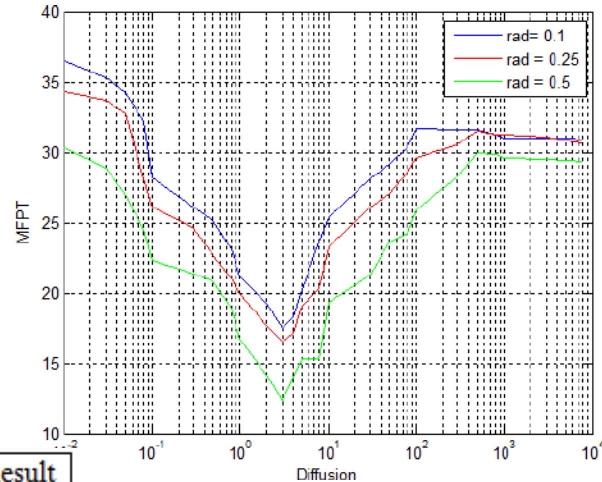
First reaction time is defined as the time when particle reside in reacting space and react. In marketing , “reaction” means viral

MFPT Without Force.

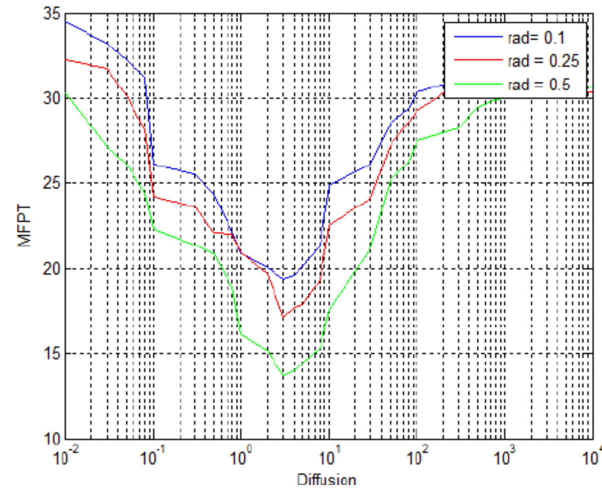


	Simulation result	Analytical Result
r=0.1	31.6949	31.32
r=0.25	5.0459	5.091
r=0.5	1.2248	1.273

MFPT With Constant Force.



MFPT With Spring Force.

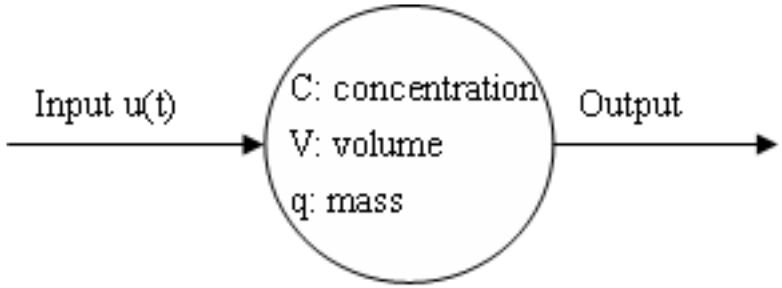


This optimal diffusion coefficient trend and value are shift by radius of reaction space. The bigger radius , the smaller optimal diffusion coefficient.

Possibility of viral is bigger when exposure network is bigger

Information Spreading (e.g Hoax spreading) : Compartment Model(health/disease)

Also famous as SIR Model

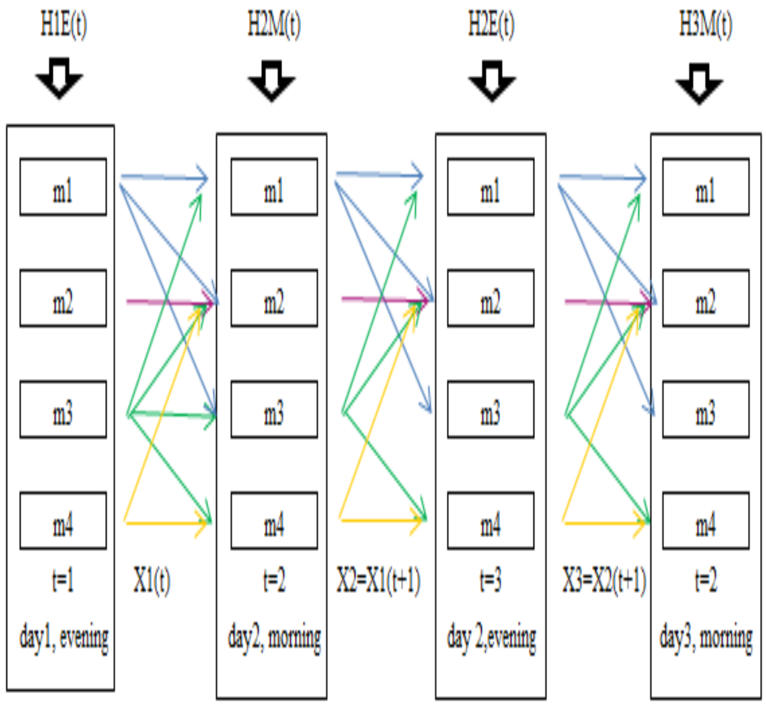


$$\frac{\mathrm{d}q}{\mathrm{d}t} = u(t) - kq$$

$$C = \frac{q}{V}$$

where *k* is the proportionality.

Combining Compartment model with network analysis



The Algorithm of Spreading

1. Initialization: - weather factor: W(t-1)

   - Dengue-contracting risk value $\overline{L} = (L_1(t), L_2(t), L_3(t), \ldots, L_n(t))$

   - Initial origin-destination vector $\overline{M}(t) = (M_1(t), M_2(t), \ldots, M_n(t))$

2. Calculate $P_n(t)$

3. Calculate Risk Area H1M(t), H2M(t), H2E(t) standard include the weather factor.

4. Input : Infected people $\overline{MI_n}(t+1)$

   If Infected people = $\overline{MI_n}(t+1) = (M_1(t+1), \ldots, M_n(t+1))$

   then $L_n(t+1) = 1$

   else $L_n(t+1) = L_n$

5. Repeat step 3 for t= t+1

6. Calculate $DH1M_m = H1M_m(t+1) - H1M_m(t)$

   $DH2E_m = H2E_m(t+1) - H2E_m(t)$

   $DH2M_m = H2M_m(t+1) - H2M_m(t)$

7. Check : if $DH1M_m = 0$, $DH2E_m = 0$, $DH2M_m = 0$

   then stop

   else select from vector $\overline{M}(t)$ to get $AI_n(t+1)$

   $AI_n(t+1)$ are Area Infected with specification

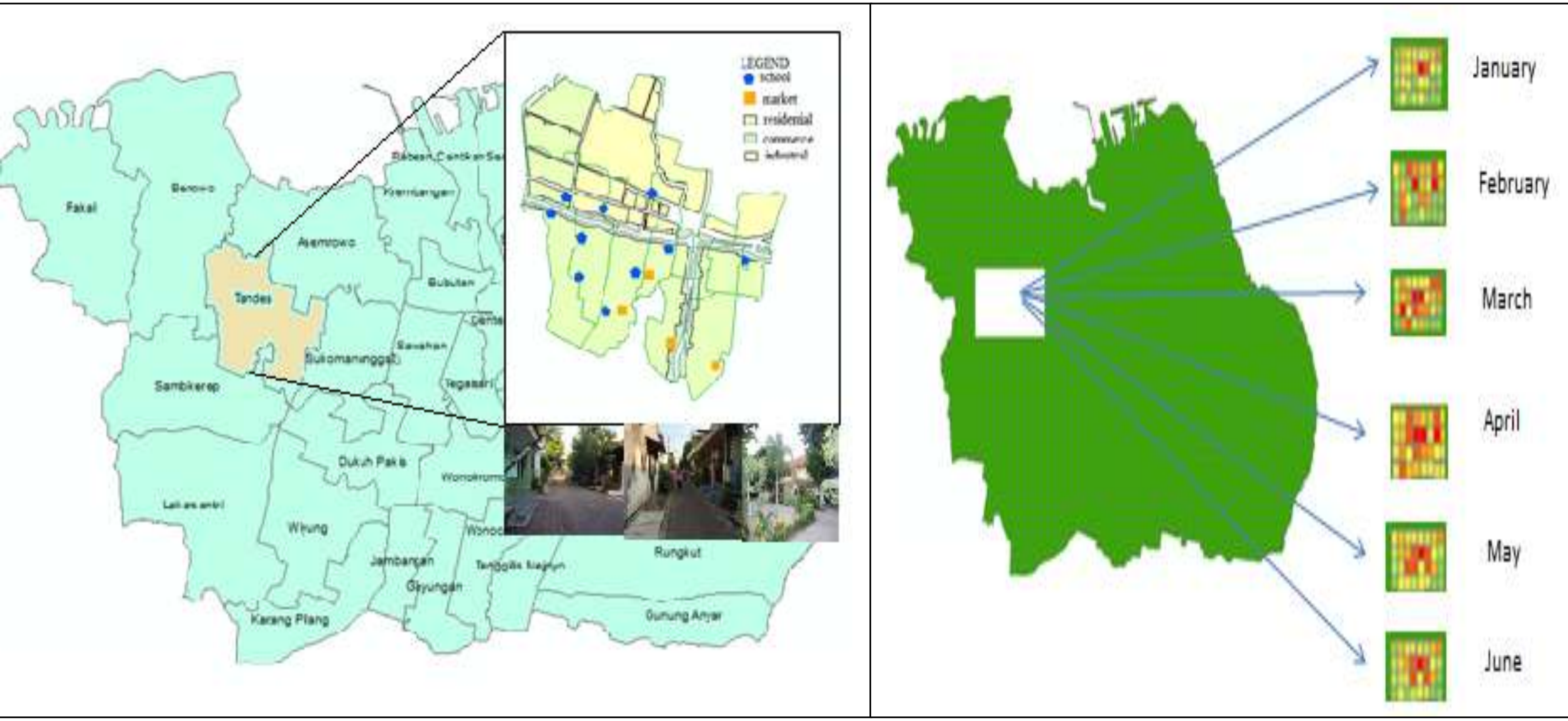
   $DH1M_m \neq 0$, $DH2E_m \neq 0$, $DH2M_m \neq 0$

8. Select next suspected area from vector $\overline{MI_n}(t+1)$ by inputting $AI_n(t+1)$

9. Update standard H1M(t), H2M(t), H2E(t) with H1M(t+1), H2M(t+1), H2E(t+1)

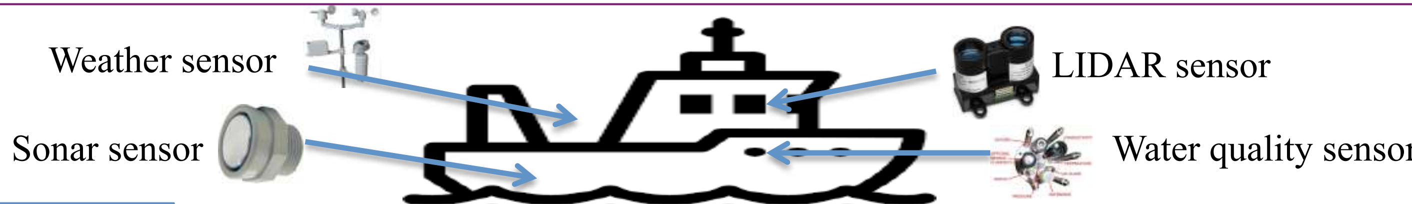
10. Repeat process for t = t+n with condition :

    H1M(t-2)= H1M(t), H2M(t-2)= H2M(t), H2E(t-2)= H2E(t)



Fakal
Benowo
Asemrowo
Krembangan
Tandes
Bubutan
Sukomanunggal
Sawahan
Sambikerep
Tegalsari
Dukuh Pakis
Wiyung
Jambangan
Gayungan
Rungkut
Karang Pilang
Gunung Anyar
LEGEND
school
market
residential
commerce
industrial
January
February
March
April
May
June

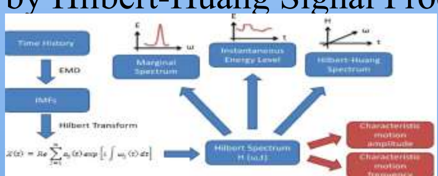
Event-Base Analysis : Hilbert Huang (Signal Processing)



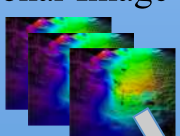
Water data

Weather data

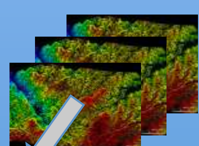
Event-base feature extrac to find the critical event in coral disease by Hilbert-Huang Signal Proc



Sonar Image

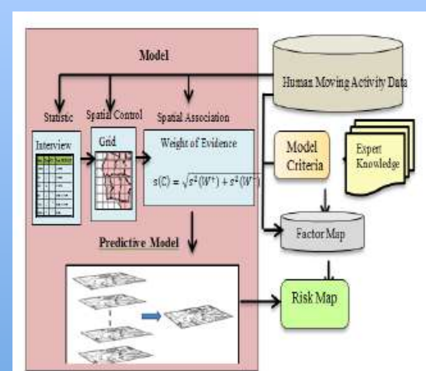


Lidar Image

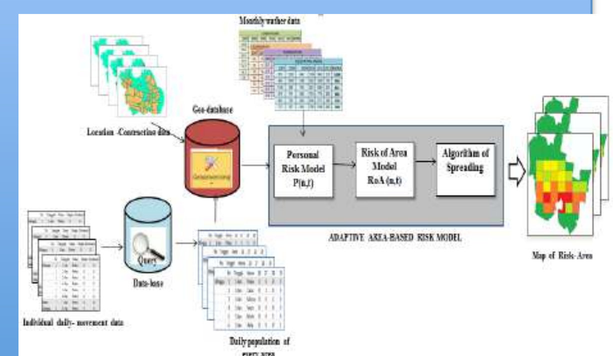


Vector Differential Image in Hilbert Space to analyze the first critical damage ima

Fusion sensing data system

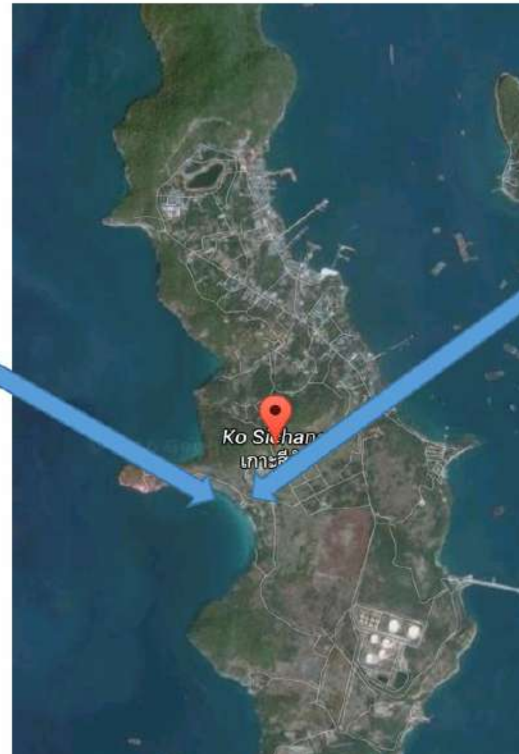


Bacterial-contraction Area Risk System

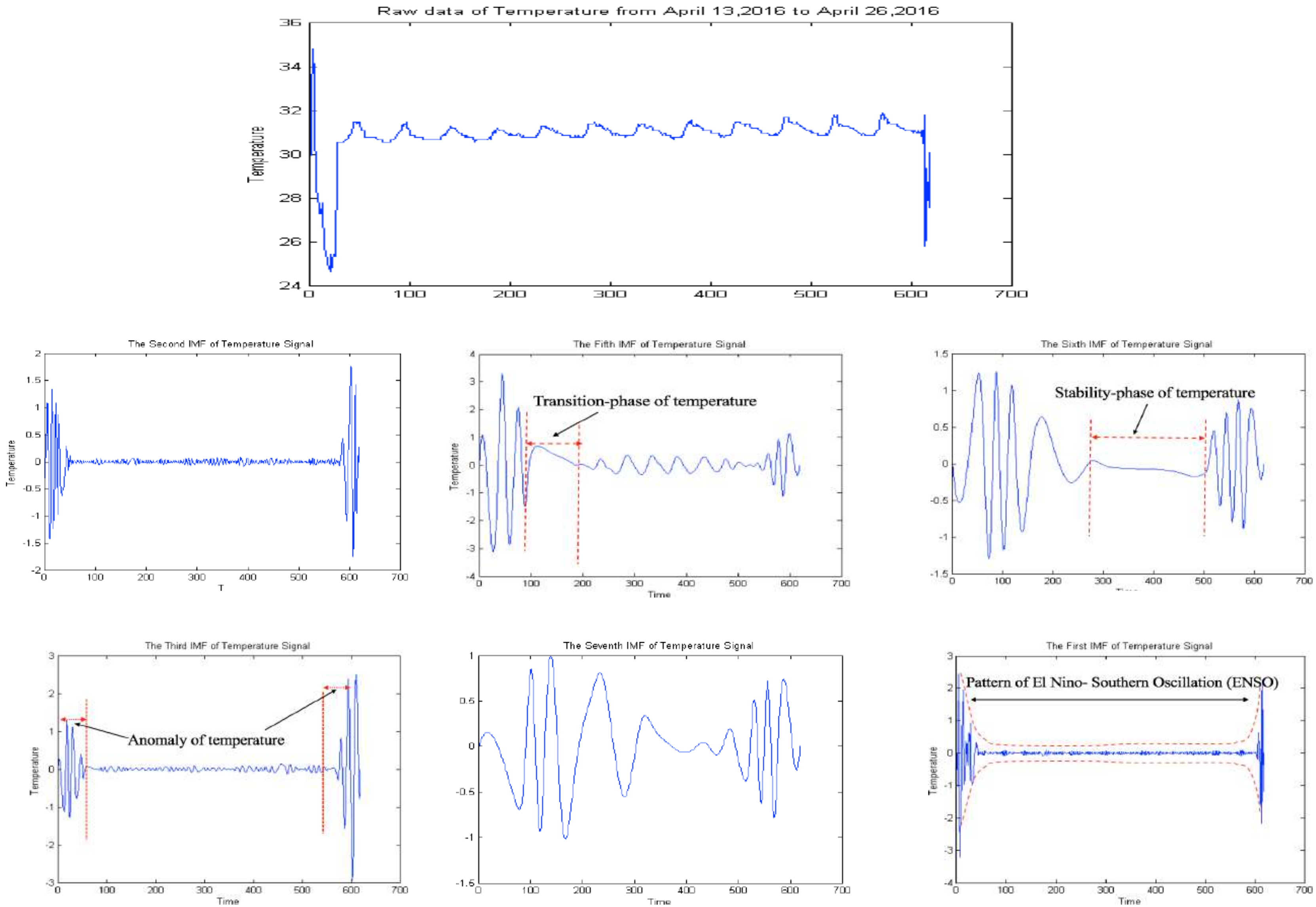


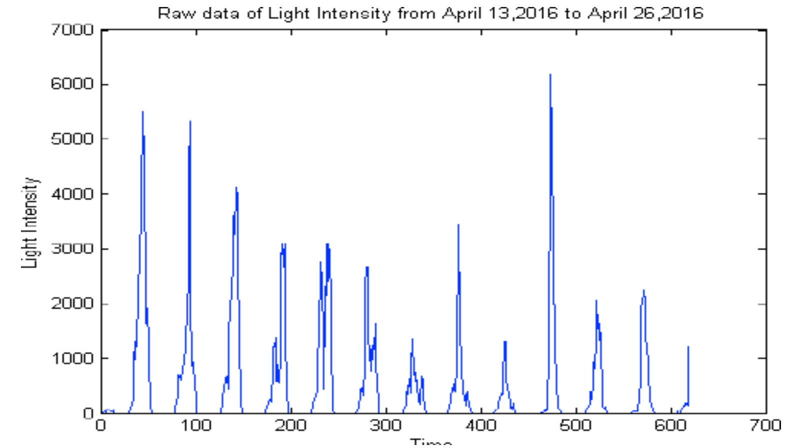
Adaptive Area-base Risk System

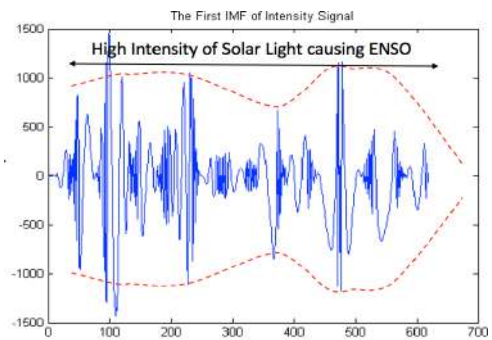
Ka Chang Area

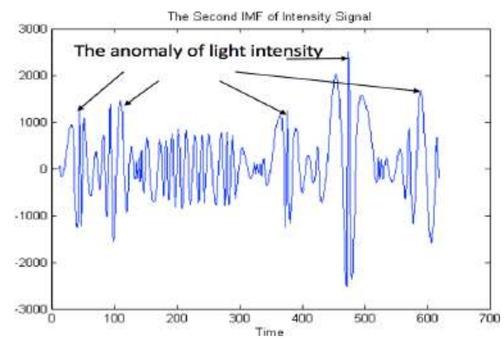


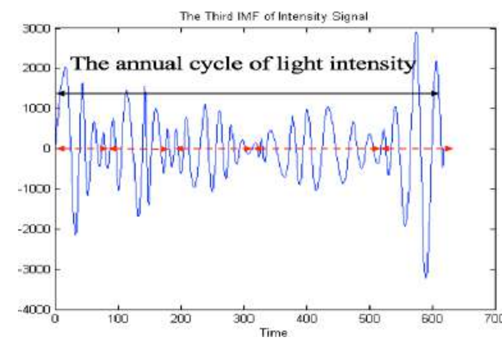
Sichang Marine Science Research and Training Station (SMaRT).

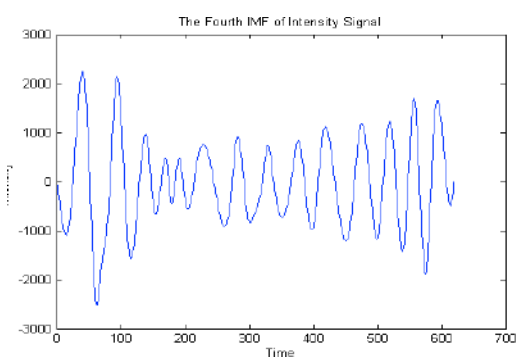

Raw data of Temperature from April 13,2016 to April 26,2016
Temperature
The Second IMF of Temperature Signal
The Fifth IMF of Temperature Signal
Transition-phase of temperature
Time
The Sixth IMF of Temperature Signal
Stability-phase of temperature
The Third IMF of Temperature Signal
Anomaly of temperature
The Seventh IMF of Temperature Signal
The First IMF of Temperature Signal
Pattern of El Nino- Southern Oscillation (ENSO)

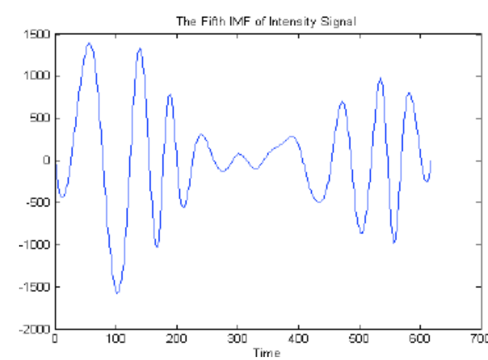

Raw data of Light Intensity from April 13,2016 to April 26,2016
Light Intensity
Time

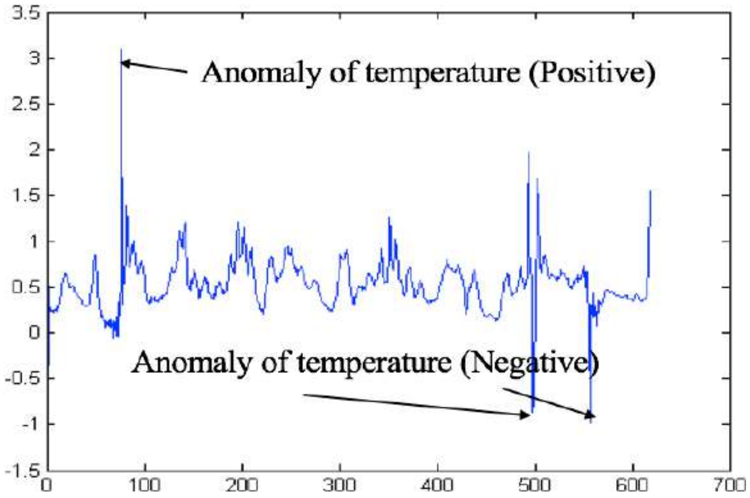

The First IMF of Intensity Signal
High Intensity of Solar Light causing ENSO

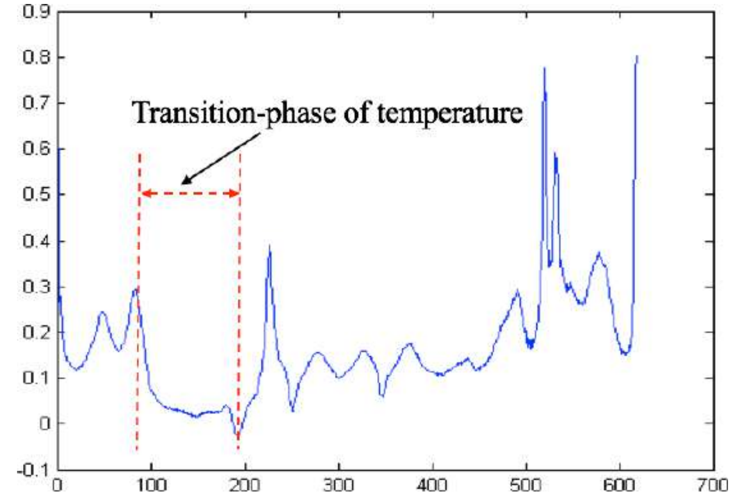

The Second IMF of Intensity Signal
The anomaly of light intensity
Time

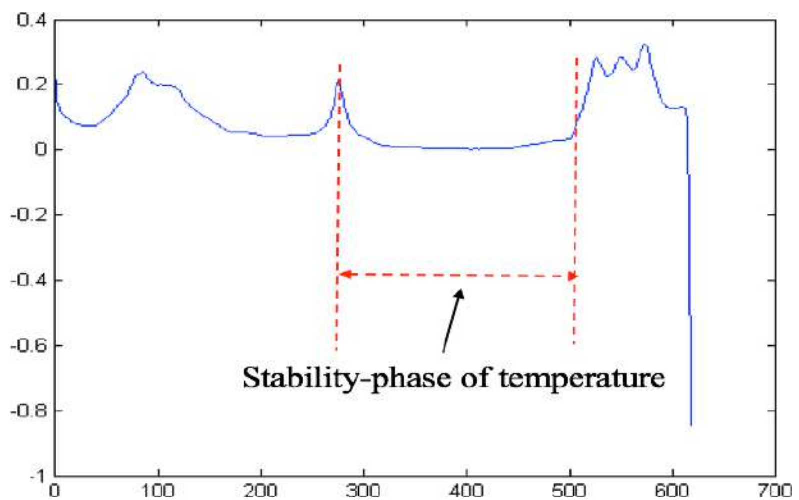

The Third IMF of Intensity Signal
The annual cycle of light intensity
Time

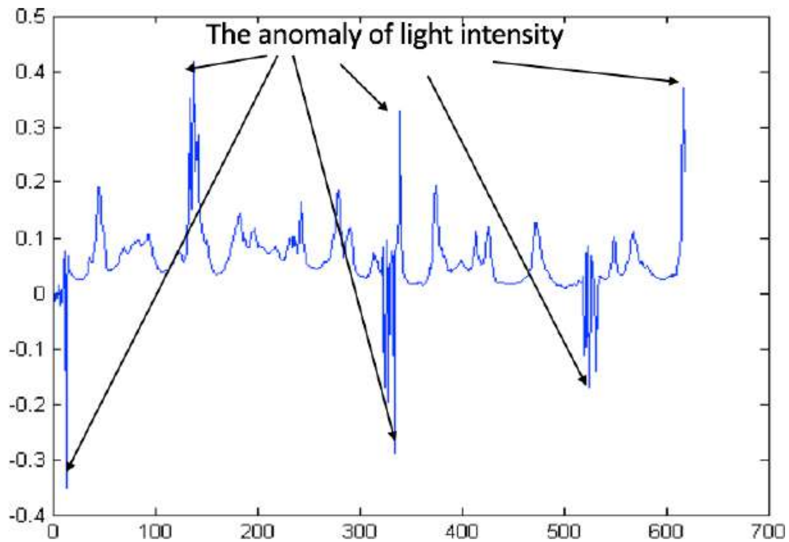

The Fourth IMF of Intensity Signal
Time

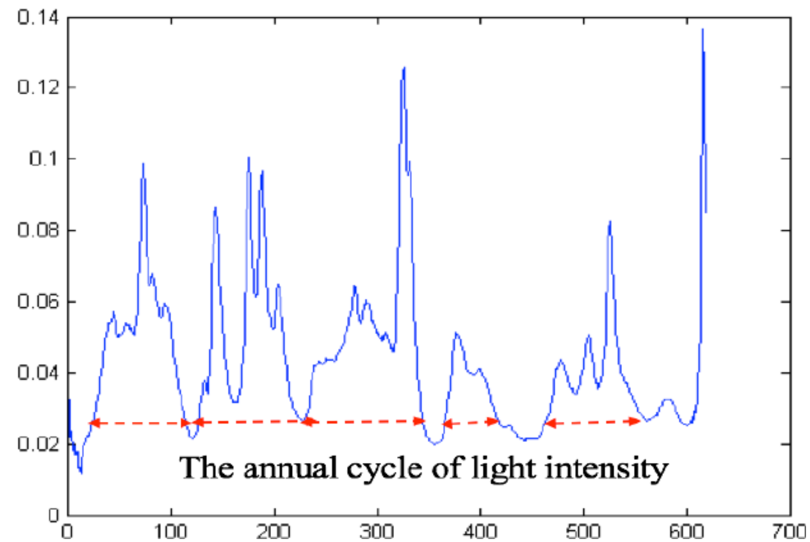

The Fifth IMF of Intensity Signal
Time


Anomaly of temperature (Positive)
Anomaly of temperature (Negative)
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
0
100
200
300
400
500
600
700


Transition-phase of temperature
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
0
100
200
300
400
500
600
700


Stability-phase of temperature
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
100
200
300
400
500
600
700

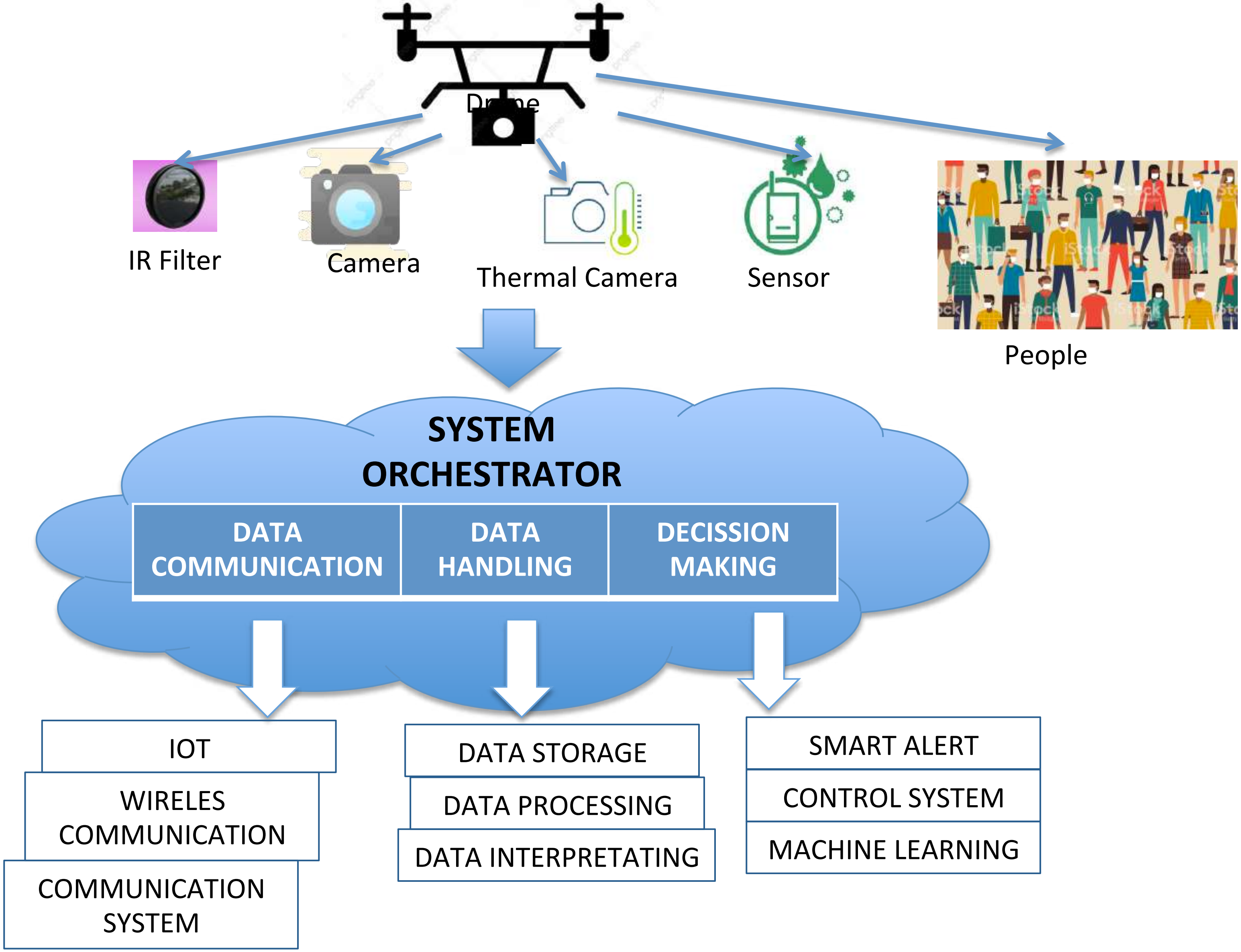

The anomaly of light intensity
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
0
100
200
300
400
500
600
700

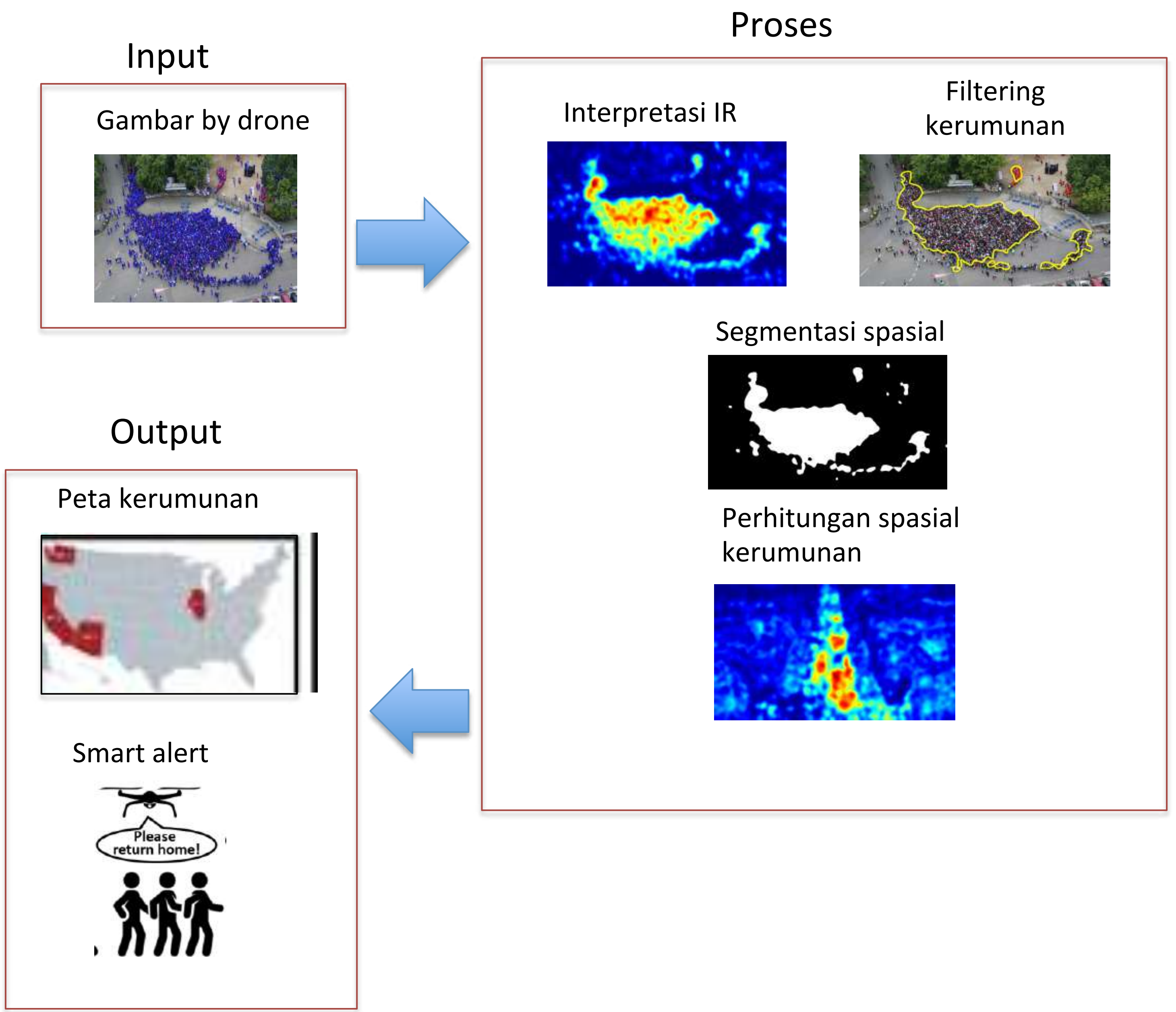

The annual cycle of light intensity
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
100
200
300
400
500
600
700

Event	Description	Physical Phenomena	Period (Time) of measurement	Frequency
ENSO	An abnormal weather pattern caused by the warming of the Pacific Ocean near the equator	Stationary hot wave	0-624	From (3) To (-3)
Temperature anomaly	Impulsive-extreme temperature	Impulsive of low or high temperature	Hot : 92 Cold : 500,576	Hot : 3 Cold : -0.72,-0.84
Transition Phase	Transition from extreme temperature to annual cycle	Temperature rising or decreasing gradually	From 94 to 100	From 0.3 to 0
Stability Phase	An annual cycle of temperature	Cyclic temperature oscillation	From 284 to 502	From 0.2 to 0

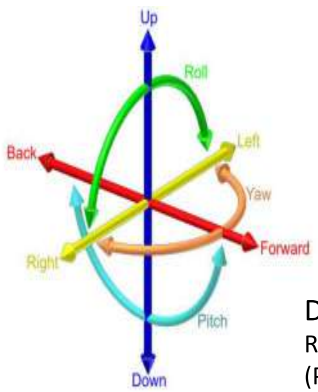
Event of Temperature	Event of Light Intensity	Incubation Time
ENSO	ENSO	122
ENSO	Light intensity anomaly	104
ENSO	Annual cycle	112
Temperature Anomaly	ENSO	86
Temperature Anomaly	Light intensity anomaly	73
Temperature Anomaly	Annual cycle	88
Transition Phase	ENSO	118
Transition Phase	Light intensity anomaly	92
Transition Phase	Annual cycle	146
Stability Phase	ENSO	110
Stability Phase	Light intensity anomaly	101
Stability Phase	Annual cycle	136

Crowd Analysis:Fluid Mechanic (Mechanical engineering)

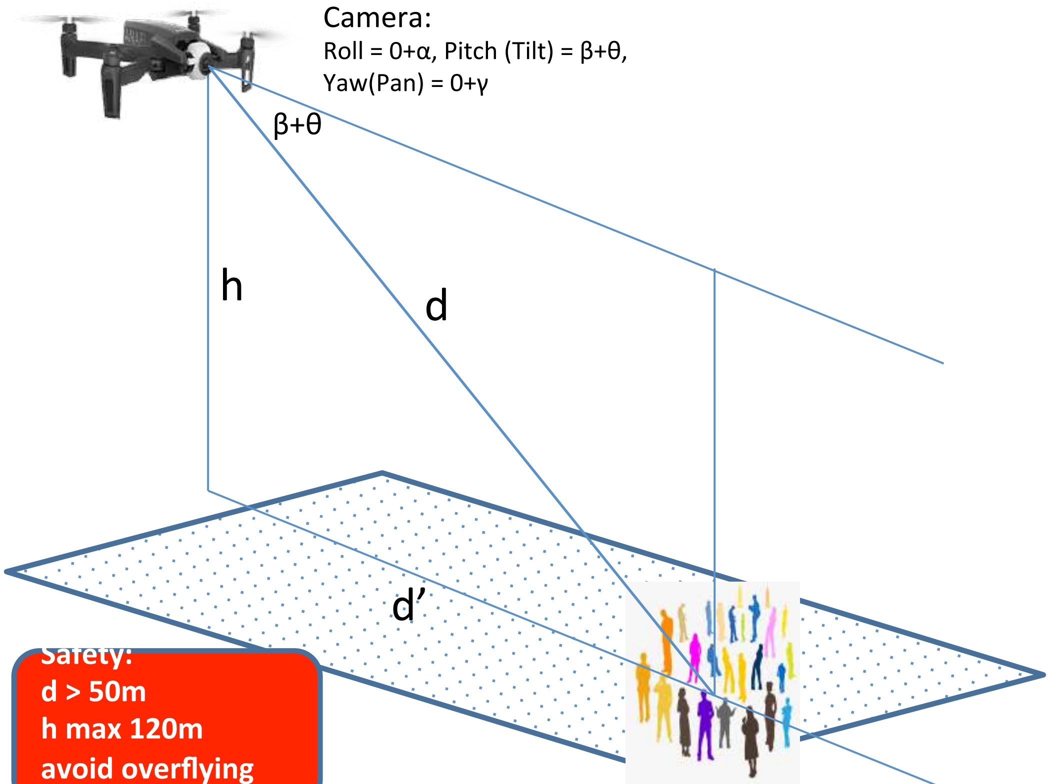



Input
Gambar by drone
Proses
Interpretasi IR
Filtering kerumunan
Segmentasi spasial
Perhitungan spasial kerumunan
Output
Peta kerumunan
Smart alert
Please return home!

Konsep Pendeteksi Kerumunan dan Pengumpulan Dataset



Drone:
Roll = α, Pitch (Tilt) = β, Yaw (Pan) = γ
Camera:
Roll = 0+α, Pitch (Tilt) = β+θ, Yaw(Pan) = 0+γ



Safety:
d > 50m
h max 120m
avoid overflying the crowd

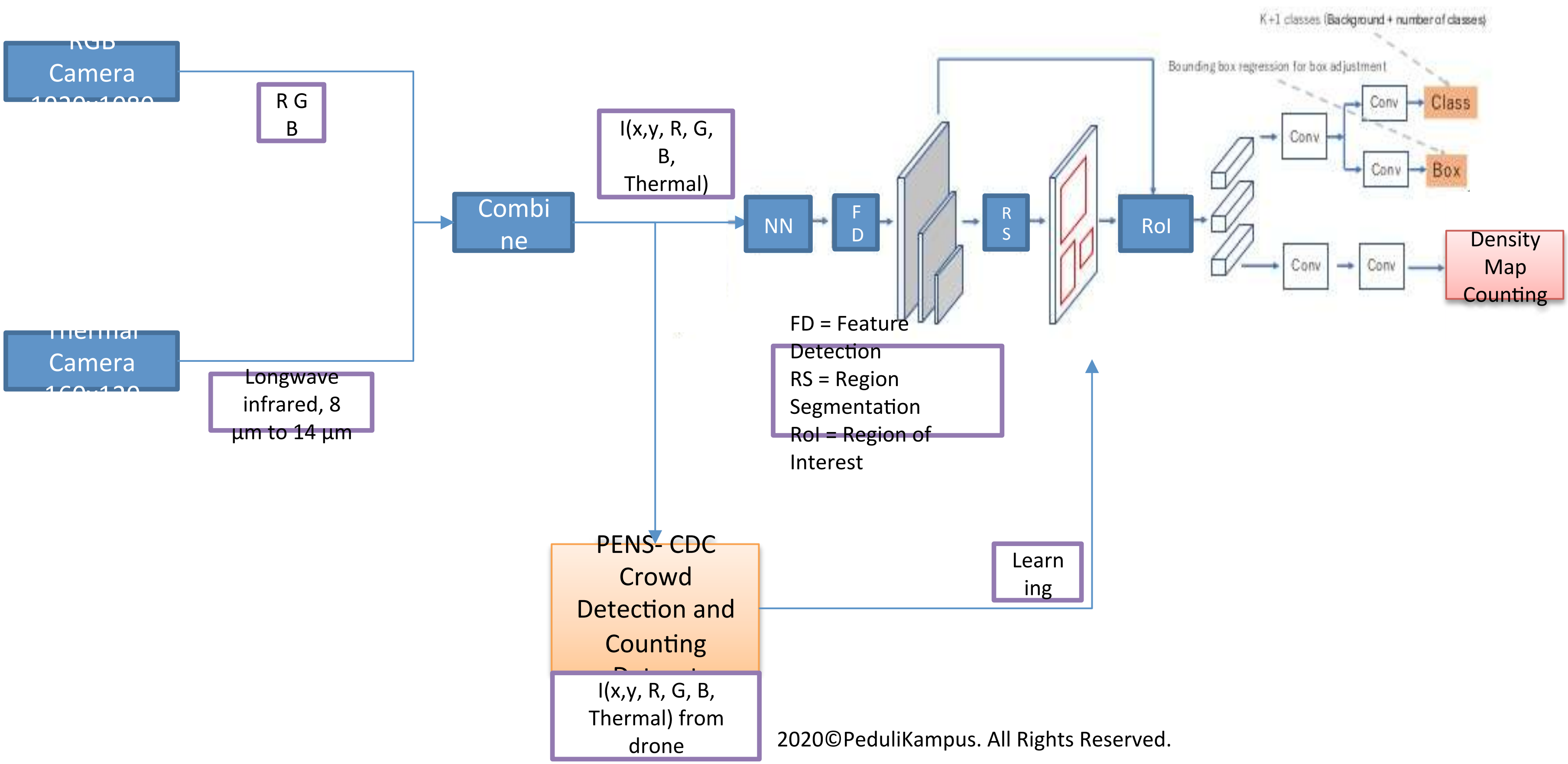
Pengambilan Dataset (Training)
α = β = γ = 0 (mendekati dan dicatat sudutnya, α lebih kecil
θ = 30 (dataset A)
θ = 45 (dataset B)
θ = 60 (dataset C)

Pengambilan Dataset (Testing)
α = β = γ = 0 (mendekati dan dicatat sudutnya , α lebih kecil
θ = 30 (dataset A)
θ = 45 (dataset B)
θ = 60 (dataset C)
β+θ = 15 s/d 75 (random)

Data diambil pada saat drone:
- Diam pada ketinggian h.
- melakukan gerakan yaw rotation (turning), kearah ± yaw
- melakukan gerakan lateral kedepan (forward), kekiri atau kekanan (sideway)
- gerakan arcs maneuver, gerakan berputar mengelilingi satu target (± 120)

θ = 30, h = 25m
θ = 45. h = 35.35m
θ = 60. h = 43.30m

Framework Crowd Detection and Counting



Crowd Detection Drone

Spectral Camera

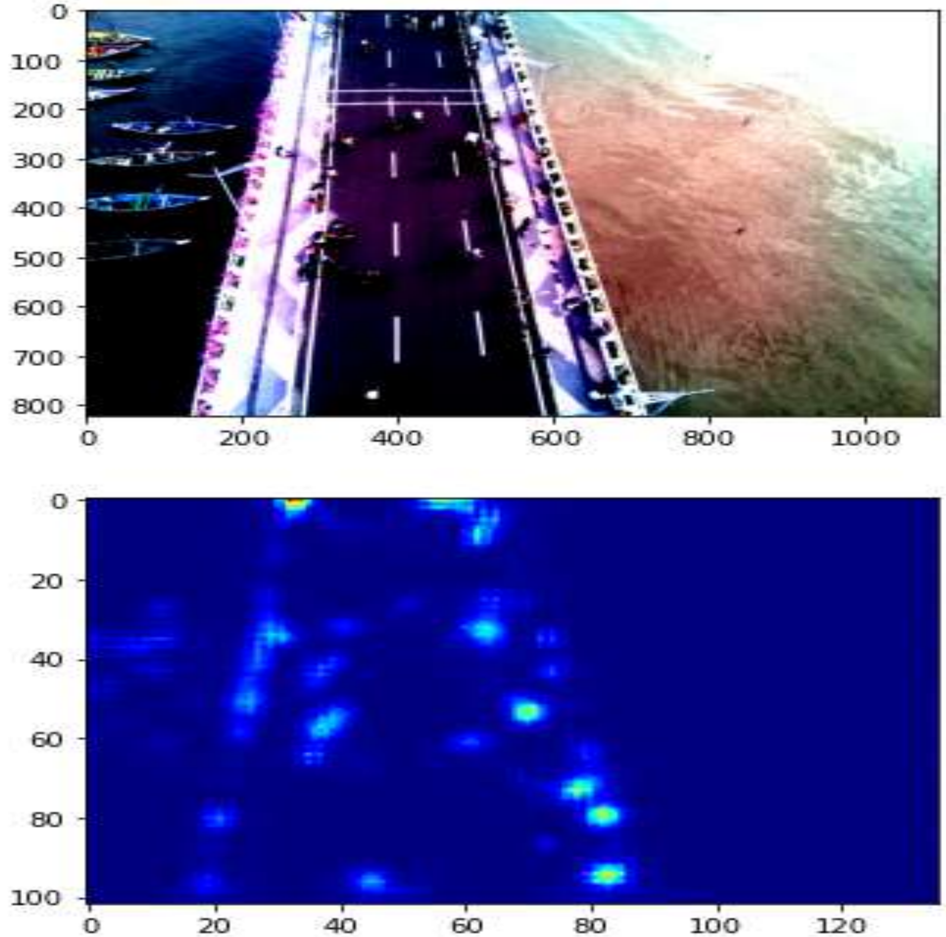
Telemetri Test

Test jarak komunikasi telemetry (merah = ground, kuning = 10 m, biru = 20 m dan hijau = 30 m)

Detection Result (Suramadu)

18.500908

Clipping input data to the valid range for im



Detection Result (Pasar Pacar Keling)

84.28146