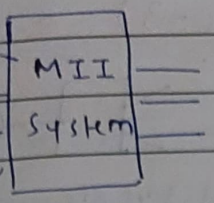
7 9th Feb, 23

MII-lec 8 (~~20th Feb 2023~~)



↳ not necessary all work at same time

* Data level fusion

↳ same modality

S1 S2

G1 G2

FE1 FE2

→ Fusion processes

* Feature level fusion

↳ similar modality

↳ speech & lip movement

noise affects both

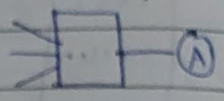
* Decision level fusion

↳ Two or more modalities are different (Face & head gesture)

noise affecting one may not affect others.

Page :
Date :

* CARE properties (lec-3)

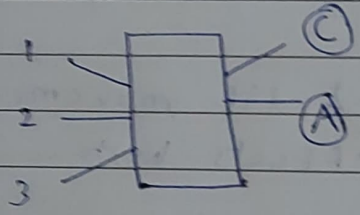


CARE = Complementarity ⇒
Assignment ⇒ 1 modality is used at a time
redundancy ⇒
equivalence ⇒

Parallel processing
+ combined (running parallel & combining all modalities)

Ex:- Put that there (speaking + pointing together to give o/p).

redundancy ⇒



if I use 1 to run the system, 2 & 3 are redundant
if I use 3, 1 & 2 = redundant

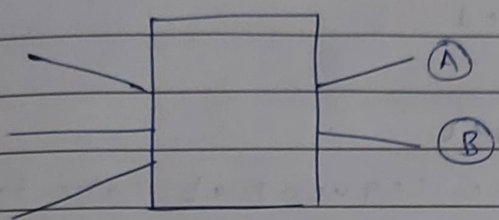
* redundancy = the modality which is not used

* Equivalence = Sequential
(Put that there ~~+ pointin~~ followed by pointing)

Vrue universsatit Brussel Stades

* CASE Properties

C = Concurrent = all commands in 1-go
A = Alternate
S = Synergetic
E = exclusive = commands at a time (1 after another) but they give diff O/P

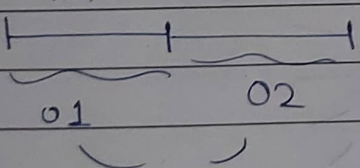


Parallel + Independent
↓
2 modalities enter at same time

* Synergestic = Parallel + Combined
  - benefit ⇒ mismatch % is less
  - most widely used

* Exclusive = sequential + independent

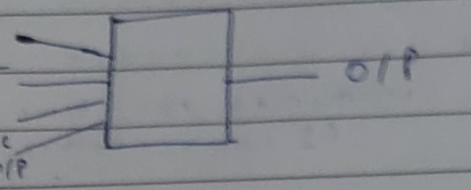
sequential modality



outputs are different

* Types of Multimodal system

(1) Fusion based (passive) unintentional
= gazing + lip movement



(2) Temporally cascaded

↳ sequential

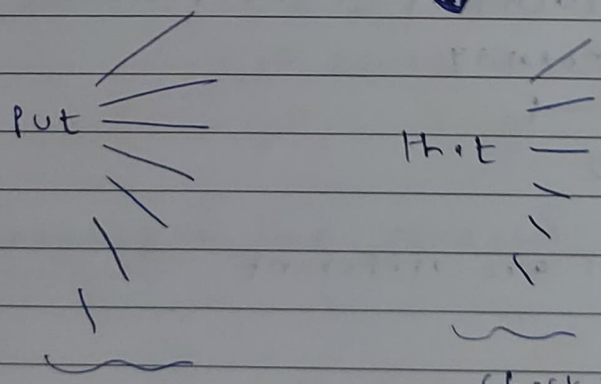
(when we use sequential data to give O/P)

Multimodal

* Fission

Speech I/P = "put that there"

After receiving, we check different versions of put



check in dict what comes after put
(This is fission process)

Check in dict & accordingly corrective measures taken

→ Detection — Colour
— 3D model = for VR
— motion based = video processing

* Hand Gesture

Data
→ ~~Hand~~ Glove

Adv: ① Accuracy ↑ (high)
② Background can be anything
③ correct O/P

Disadv: ① extra equipment
② wired which means we are constrained to 1 place (Majority)
③ not a natural way of commn.

→ coloured marker ⇒ 6th sense (P. Mistry)
↳ R, G & B markers on fingers used to control something

Adv: wireless, multiple Applications
↳ Accuracy is fine

Disadv:
↳ ① Wear the device at all times
② Similar coloured objects at background (not able to recognize) → not able to detect or track (Overlapping)

→ vision based ⇒ use of only hand and camera
↳ Adv: natural way of commn, wireless, Accuracy is fine

Disadv: Challenges =

* Skin colour ⇒ Challenges

Motion ⟶ GMM ⟶ To detect the initial object

Background subtraction

Frame differencing (when B/g is constant)

① 2 3 4 - -

If this frame no hand, then it is b.g.

2-1
3-1
4-1
5-1

} wherever motion, detection of hand

[1 2 3] 4 - - - 20

[m] ⇒ detection of hand

2-1
3-1 (diff of frames to find hand)
4-1
5-1

if slow moving of hand

→ "Illumination variation is disadvantage"
(lighting conditions)

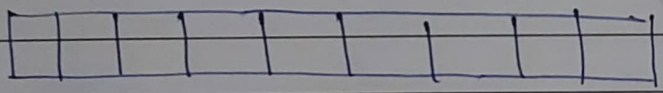
MII-Lec8 (14th Feb, 2023)

* Hand Gesture
  - Detection
  - Tracking (To locate the position of hand)
  - Recognition (meaning of the gesture)

* Limitation of frame differencing
  - B/g should be stationary

* B/g subtraction
  - The first frame should be b/g
  - Challenge = Illumination variation (A slight change in b/g)

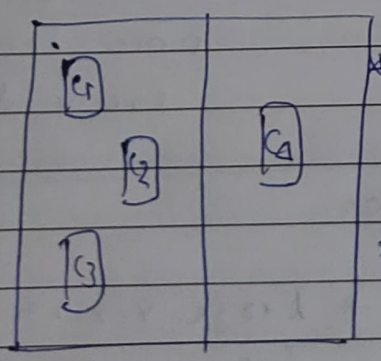
* GMM (Gaussian mixture model) → identify objects present (Training phase)
  - Test Phase



→ limitation
Training of other things can be done (like people in case of car-detection)

This can be set by ourselves

method 1:



* For first 30-seconds, training phase
* later it calculates similarity matrix to detect objects

3D model = not reqd.

Page:
Date:

before hand

Method 2: Provide dictionary & then later detect the objects

→ TRACKING

* Colour based = Marker model

* Colour based tracking
(A colour based approach)

1) Coloured marker approach

2) Skin colour object (Vision-based)

RGB → HSV (HUE-saturation-value)

YC_bC_r

$-\leq H \leq -$
$-\leq S \leq -$
$-\leq V \leq -$

Skin-colour

min & max range decided from detection of face

since face is easily detected

major disadvantages:

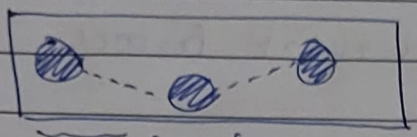
Colour overlapping + large variations (if multiple persons)

+ illumination variations

Camlin

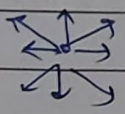
to predict the next activity

* Probabilistic approach → tracking Algo. (study about this)
  - Kalman filter (an example of prob. based approach)

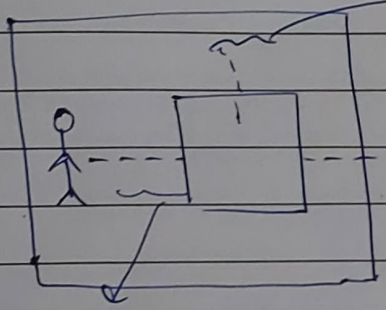


first few frames, check where object is moving

gives probability where will object move in next frame (what is next state of object)



} prob. of 8 directions



but if the changes direction, detection wrong

} if it comes to this state, detected properly

here also detected properly

→ most widely used

* Appearance based approach
  - KLT, Camshift
    - ↓ shape + colour
    - Kenneth, Lucas & Thompson

→ not ~~work~~ work well for longer videos

* KLT ⇒ feature based technique

→ Features detected from detected object & use these features in subsequent frames.

50 features extracted

edges, lines, corners

limitation = if lot of illumination variation