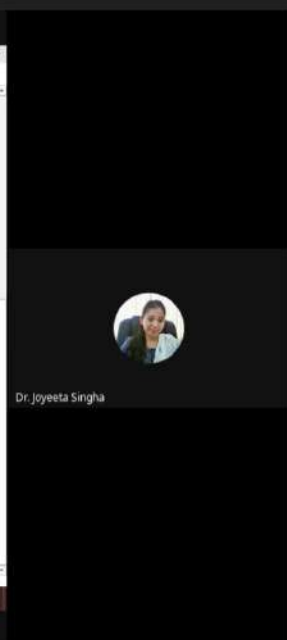


Human

5. Gustatory → taste
6. Vestibular → Balance.

Sense	Organs	Modality	Sensor
Vision	Eyes	Visual	Camera
Hearing	Ears	Aud	Microphone
Touch	Skin	Hap	Touchscreen, gloves, temp sensor
olfact.	Nose	olf	-
Taste	Tongue	Gusta	-
Balance	equilibrium org.	Vestib.	-




Dr. Joyeeta Singha


Human

6. Vestibular → Balance.

Sense	Organs	Modality	Sensor	Example / Appli
Vision	Eyes	Visual	Camera	BD just face recog
Hearing	Ears	Av	Microphone	Speech recogn
Touch	Skin	Hap	Touchscreen, gloves, temp sensor	Posture/Motion recognition
olfact.	Nose	olf	-	-
Taste	Tongue	Gusta	-	-
Balance	equilibrium	Vesti.	-	-



Dr. Joyeeta Singha




Modality :- Capability of the system to mimic such human-sensory channel.

2. Multimedia & Multimodality.

TV	Movie hall	Smart phone	PC/Desktop
----	------------	-------------	------------

Multimedia :- System with more than one way to transfer information from the system to user.





Both X

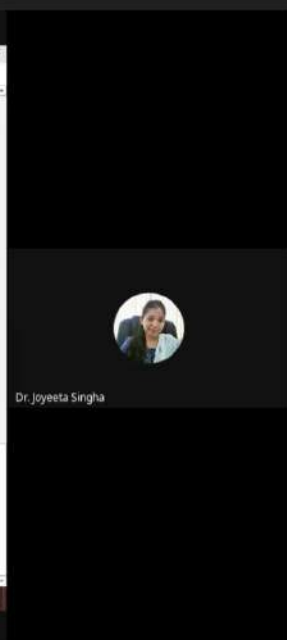
Alternate definition :-

→ Multimodal system

↳ A system that provides more than one user i/p mode to interact with it.

```
graph LR; ip1 --> System; ip2 --> System; System --> Out;
```

Some



Dr. Joyeeta Singha

aq5-xoco-huq (2021-01-09 at 01:39 GMT-8)

2nd modified definition

Multimedia system :- Interactive system that provides information via several output channels (sound + graphics)

Multimodal system :- process that is more combined user input modes (such as speech, touch, gesture, head, body...)

Diagram: A box labeled 'S+T' with 'i/p' below it has an arrow pointing to a box labeled 'S+T' with 'o/p' below it. Below the output box, it says 'that many no. of mode'.

Subtitles/closed captions (c)

24:58 / 50:01

MIL Notes For Quiz

File Edit View

2) Movie hall, smartphone, PC = can be both (but not possible in real world to be both)

Alternate Definition :

Multimodal system = A system that provides more than one user input mode to interact with it.

Ex :

- 1) Smartphone is a multimodal system - since it has a touchscreen (for touch input) and microphone (for speech input).
- 2) Robot = can be a multimodal system , can be operated with gesture, speech etc.

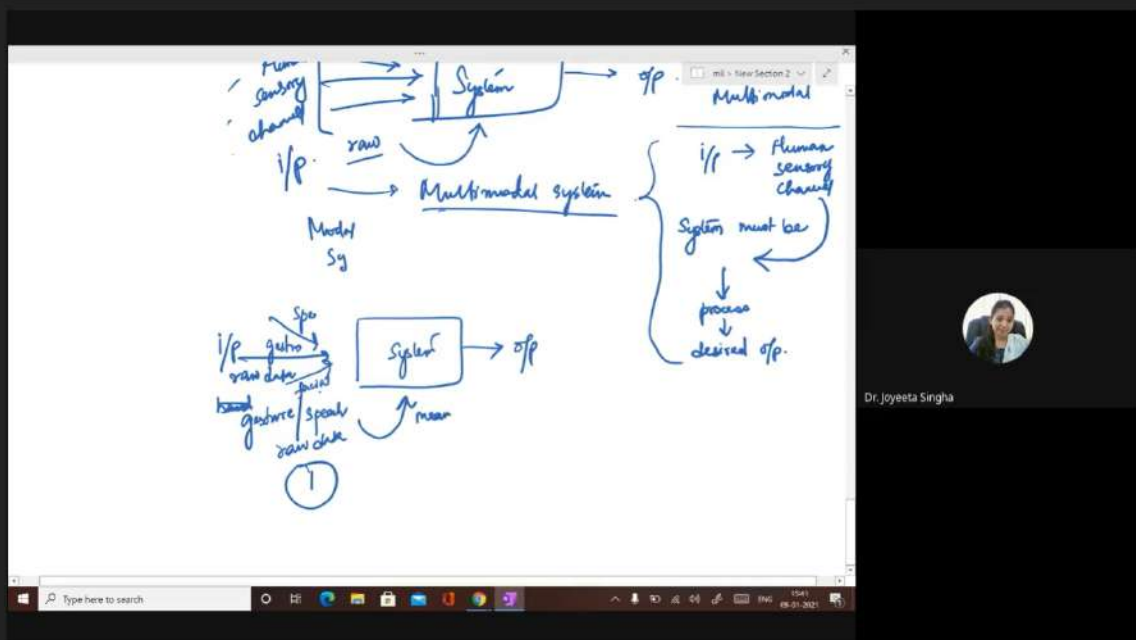
2nd modified definition :

A multimedia system is an interactive system that provides information via several output (output can be sound , graphics etc.) channels.

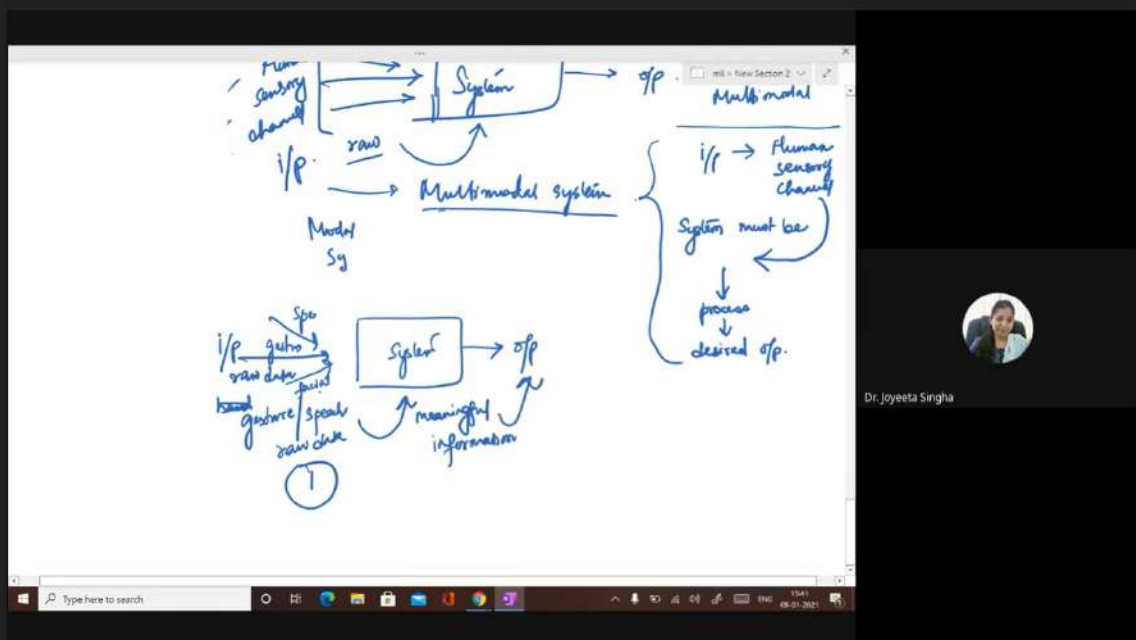
A multimodal system process two or more combined user input modes (such as speech, touch, gesture, head, body) and the output gives that many no. of modes which means that : If Speech + touch at input side, then output side also contains speech + touch.

Ln 52, Col 1 130% Windows (CRLF) UTF-8

29°C Haze 12:47 PM 2/23/2023



Dr. Joyeeta Singha



Dr. Joyeeta Singha

MIT 2023

aqx-xoco-huq (2021-01-09)

drive.google.com/file/d/1t841380_0jm-y4_eTcaW0Zp1op7suE6/view

aqx-xoco-huq (2021-01-09 at 01:39 GMT-8)

Press [Esc] to exit full screen

MIT Museum Logo

Put That There

November 2, 1979

The Architecture Machine

© 1979 MIT

Dr. Joyeeta Singha

44:31 / 50:01

29°C Haze

Search

ENG IN

1:03 PM 2/23/2023

hsu-xyby-kmy (2021-01-11 at 02:48 GMT-8)

1. Active- I/p mode
2. Passive- I/p "
3. Virtual environment

1.
i/p (intentionally)
[1, 2, 3, 4]
[]

User can intentionally give information to the system e.g. Command | gesture

Lecture - 2

File Edit View

Lecture - 2

Multimodal systems are subset of Multimedia systems.

Domains of multimodal systems :

1) Active I/P mode
2) Passive I/P mode
3) Virtual Environment = mostly used in gaming applications

1) Active I/P mode = user intentionally gives information to the system. Ex : Command, gesture

Ln 14, Col 1 130% Windows (CRLF) UTF-8

hsu-xyby-kmy (2021-01-11 at 02:48 GMT-8)

Active I/P mode
 Intentionally → desired O/P.

2. Passive I/P mode
 → All kinds of information that user does not produce intentionally.
 e.g. emotions/gazing something / ...

System → O/P.

(Not intentionally provided by the user)

Lecture - 2

"Multimodal systems are subset of Multimedia systems."

Domains of multimodal systems :

- 1) Active I/P mode
- 2) Passive I/P mode
- 3) Virtual Environment = mostly used in gaming applications

1) Active I/P mode = user intentionally gives information to the system. Ex : Command, gesture
 Ex : A Dialog system with different input modes.

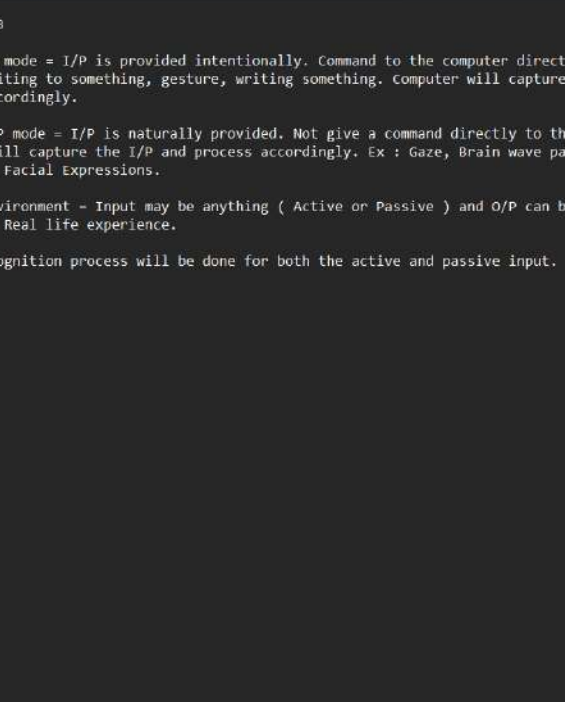
Fixed Input = to give desired O/P

2) Passive I/P mode = All kinds of information that user does not produce intentionally. No fixed pattern at input end.
 Ex : Emotions, gazing at something

Ln 21, Col 1 | 130% | Windows (CRLF) | UTF-8

The image contains several handwritten diagrams illustrating computer architectures:

- Top Left:** A diagram showing a **VP** (User) sending a **request** to a **System** (represented by a box). The system outputs **data**. A note says: "Sequentially processed, connected to the computer directly".
- Top Right:** A diagram showing a **VP** sending a **request** to a **Control** unit, which then sends a **request** to a **System** (represented by a box). A note says: "naturally processed".
- Middle:** A diagram showing a **VP** sending a **request** to a **System** (represented by a box). The system outputs **data**. A note says: "not the user give a command directly to the computer".
- Bottom Left:** A diagram showing a **VP** sending a **request** to a **System** (represented by a box). The system outputs **data**. A note says: "Perform this task...".
- Bottom Center:** A diagram showing a **VP** sending a **request** to a **System** (represented by a box). The system outputs **data**. A note says: "gave, broken many patterns, lip movement, facial expression".
- Bottom Right:** A diagram showing a **VP** sending a **request** to a **System** (represented by a box). The system outputs **data**. A note says: "process / recog".



Lecture - 3

Active I/P mode = I/P is provided intentionally. Command to the computer directly. Ex : Speech, Pointing to something, gesture, writing something. Computer will capture the I/P and process accordingly.

Passive I/P mode = I/P is naturally provided. Not give a command directly to the computer. Computer will capture the I/P and process accordingly. Ex : Gaze, Brain wave patterns, Lip movements, Facial Expressions.

Virtual Environment - Input may be anything (Active or Passive) and O/P can be provided virtually. Real life experience.

NOTE : Recognition process will be done for both the active and passive input.

Ln 9, Col 79 100% Windows (CRLF) UTF-8

3:48 PM 2/23/2023

Problem description [edit]

The problem to be solved is detection of faces in an image. A human can do this easily, but a computer needs precise instructions and constraints. To make the task most manageable, Viola-Jones requires full-view, frontal upright faces. Thus in order to be detected, the entire face must point towards the camera and should not be tilted to either side. While it seems these constraints could diminish the algorithm's utility somewhat, because the detection step is most often followed by a recognition step, in practice these limits on pose are quite acceptable.

Components of the framework [edit]

Feature types and evaluation [edit]

The characteristics of Viola-Jones algorithm which make it a good detection algorithm are:

- Robust – very high detection rate (true-positive rate) & very low false-positive rate always.
- Real time – For practical applications, at least 2 frames per second must be processed.
- Face detection only (not recognition) – The goal is to distinguish faces from non-faces (detection is the first step in the recognition process).

The algorithm has four stages:

1. [Face Feature Selection](#)
2. [Creating an Integral Image](#)
3. [Adaboost Training](#)
4. [Cascading Classifiers](#)

The features sought by the detection framework universally involve the sums of image pixels within rectangular areas. As such, they bear some resemblance to Haar basis functions, which have been used previously in the realm of image-based object detection [2]. However, since the features used by Viola and Jones are only on more than one rectangular area, they are generally more complex. The figure on the right illustrates the four different types of features used in the framework. The value of any given feature is the sum of the pixels within clear rectangles subtracted from the sum of the pixels within shaded rectangles. Rectangular features of this sort are primitive when compared to alternatives such as steerable filters. Although they are sensitive to vertical and horizontal features, their feedback is considerably coarser.

Dr. Joyeeta Singha

drive.google.com/file/d/1dvSho5_x3_x6i_aA/As2lZKQk4RkFG1h/view

uwj-umus-qbx (2021-02-10 at 01:47 GMT-8)

vision.comml.edu/in5/sup-contents/topics/edu/2014/08/psm02.pdf

112058-1.1.25

1 / 28 156%

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

Face image acquisition cost imply that vision systems can be deployed in desktop and d systems [111], [112], [113]. The rapidly expand- ch in face processing is based on the premise that on about a user's identity, state, and intent can be from images, and that computers can then react- ly, e.g., by observing a person's facial expression. t five years, face and facial expression recognition racted much attention though they have been for more than 20 years by psychophysicists, ntists, and engineers. Many research demonstra- l commercial applications have been developed e efforts. A first step of any face processing system ng the locations in images where faces are present. , face detection from a single image is a challen- because of variability in scale, location, orientation , rotated), and pose (frontal, profile). Facial ex, occlusion, and lighting conditions also change ll appearance of faces.

Yang is with Honda Fundamental Research Labs, 800 California koutem View, CA 94041. E-mail: anyang@hrla.com. gnan is with the Department of Computer Science and Beckman , University of Illinois at Urbana-Champaign, Urbana, IL 61801. anyang@hrla.com

- **Pose.** The images of a face vary due to the relative camera-face pose (frontal, 45 degree, profile, upside down), and some facial features such as an eye or the nose may become partially or wholly occluded.
- **Presence or absence of structural components.** Facial features such as beards, mustaches, and glasses may or may not be present and there is a great deal of variability among these components including shape, color, and size.
- **Facial expression.** The appearance of faces are directly affected by a person's facial expression.
- **Occlusion.** Faces may be partially occluded by other objects in an image with a group of people, some faces may partially occlude other faces.
- **Image orientation.** Face images directly vary for different rotations about the camera's optical axis.
- **Imaging conditions.** When the image is formed, factors such as lighting (spectra, source distribution and intensity) and camera characteristics (sensor response, lenses) affect the appearance of a face.

There are many closely related problems of face detection. Face localization aims to determine the image position of a single face; this is a simplified detection problem with the assumption that an input image contains

Dr. Joyeeta Singha