



VIDEO ANALYTICS MODULE # 1 : VIDEO ANALYTICS

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

DL Team, BITS Pilani

The instructor is gratefully acknowledging
the authors who made their course
materials freely available online.

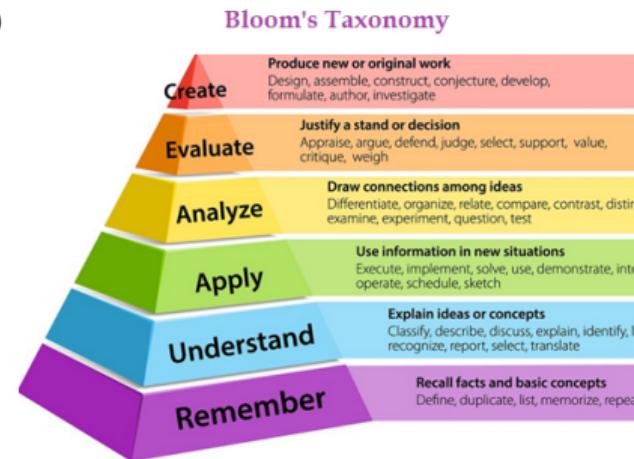
This deck is prepared by Seetha Parameswaran.

TABLE OF CONTENTS

- 1 COURSE LOGISTICS
- 2 MODULE 1
- 3 DIGITAL VIDEO
- 4 VIDEO ANALYTICS
- 5 APPLICATIONS OF VIDEO ANALYTICS
- 6 ANALOG VIDEO FORMATS
- 7 SPATIO TEMPORAL SAMPLING STRUCTURE

COURSE OBJECTIVES

- CO1** Students should gain a working knowledge of video analytics.
- CO2** Students should be familiar with various building block algorithms in video analytics, including Image and Video processing and Deep Learning with emphasis on the algorithm building blocks.
- CO3** Students should create at least one end-user application.



COURSE HANDOUT = WHAT WE LEARN...

- ① Video Analytics(4 hrs)
- ② Motion Detection and Estimation (6 hrs) (T1 Ch3)
- ③ Video Enhancement and Restoration (4 hrs) (T1 Ch 4)
- ④ Video Segmentation (6 hrs) (T1 Ch 6)
- ⑤ Motion Tracking in Video (6 hrs) (T1 Ch 7)
- ⑥ Video Indexing, Summarization, Browsing, and Retrieval (4 hrs) (T1 Ch 15)
- ⑦ Discussion and advanced topics and applications (2 hrs)

TEXT AND REFERENCE BOOKS

TEXT Books

- T1 | Bovik, Alan C. The essential guide to video processing. Academic Press, 2009.

REFERENCE Books

- R1 | Tekalp, A. Murat. Digital video processing. Prentice Hall Press, 2015.
- R2 | Bovik, Alan C. Handbook of image and video processing. Academic press, 2010.

LAB SESSIONS

- ① Python Libraries – OpenCV, Keras and Tensorflow
- ② Matlab – Image processing, Video processing, ML / DL toolboxes

L1 Reading video and Displaying frames from video

L2 Video pre-processing

L3 Motion Detection and Estimation

L4 Video Enhancement and Restoration

L5 Video Segmentation

L6 Motion Tracking in Video and Kalman filtering for object tracking

L7 Video Indexing and Summarization

L8 Applications

LMS

Most relevant and up to date info on Canvas

- Handout
- Schedule for Webinar, Quiz, and Assignments.
- Session Slide Decks
- Demo Lab Sheets
- Quiz-I, Quiz-II
- Assignment-I, Assignment-II

The video recording will be available in Microsoft Teams.

EVALUATION SCHEDULE

No	Name	Type	Duration	Weight	Day, Date, Time
EC1	Quiz I	Online	0.5 hr	5 %	
	Quiz II	Online	0.5 hr	5 %	
	Assignment I	Online	4 weeks	15 %	
	Assignment II	Online	4 weeks	15 %	
EC2	Mid-sem Regular	Closed book	2 hrs	30 %	
EC3	Compre-sem Regular	Open book	2.5 hrs	30 %	

TABLE OF CONTENTS

- 1 COURSE LOGISTICS
- 2 MODULE 1
- 3 DIGITAL VIDEO
- 4 VIDEO ANALYTICS
- 5 APPLICATIONS OF VIDEO ANALYTICS
- 6 ANALOG VIDEO FORMATS
- 7 SPATIO TEMPORAL SAMPLING STRUCTURE

MODULE TOPICS....

- Introduction to Video Analytics
- Applications of Video Analytics
- Digital Video (T1 Ch 1, R1 Ch 1.2)
- Spatio temporal sampling structures (T1 Ch 2)

TABLE OF CONTENTS

- 1 COURSE LOGISTICS
- 2 MODULE 1
- 3 DIGITAL VIDEO
- 4 VIDEO ANALYTICS
- 5 APPLICATIONS OF VIDEO ANALYTICS
- 6 ANALOG VIDEO FORMATS
- 7 SPATIO TEMPORAL SAMPLING STRUCTURE

VIDEO IS EVERYWHERE !!!

Digital video applications

- Visual communication
- Video teleconferencing
- Medical video, where dynamic processes in the human body, such as the motion of the heart, can be viewed live.
- Video telephony
- Dynamic scientific visualization
- Multimedia video, where video is combined with graphics, speech, and other sensor modalities
- Video instruction
- Digital cinema (Shrek)

WHY DIGITAL VIDEO?

- Faster sensors and recording devices makes it easier to acquire and analyze digital video data sets. (Handheld digital cameras)
- Software available for simple film editing.
- HD DVD (Blu-ray), HDTV standards
- Easier wireless Internet access and mobile bandwidths (4G, 5G and beyond)

DIGITAL VIDEO

- Digital video comprises a series of digital images displayed in rapid succession. These images are called **frames**.
- Video is **dynamic** – the visual content evolves with time and generally contains moving and/or changing objects.
- Richer information
- Richness in Video results in data glut.

DIGITAL VIDEO

- Multidimensional signals
- A function of three dimensions
 - ▶ Spatial – two dimensions - x and y
 - ▶ Temporal – one dimension - time t
- Continuous signal – $I(x, y, t)$
- Discrete signal – $I(m, n, t_k)$

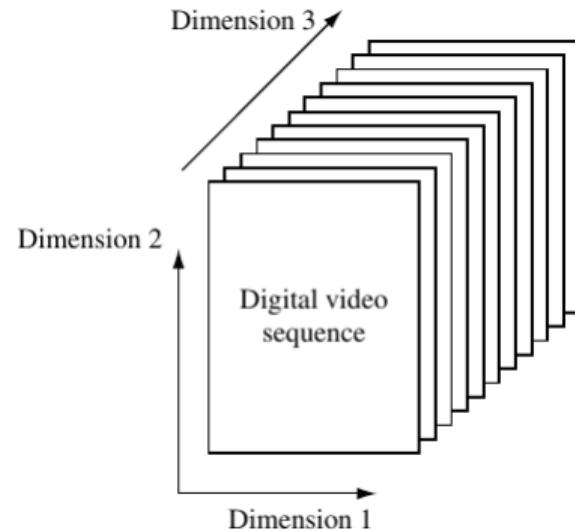


FIGURE: Digital Video

CHARACTERISTICS OF DIGITAL VIDEO

FRAME RATE The rate at which frames are displayed. Measured in frames per second (FPS).

COLOR DEPTH, OR BIT DEPTH OR RESOLUTION Fixed number of bits of a particular color channel, where the information of the color is stored within the image. 8-bit captures 256 (2^8) levels per channel, and 10-bit captures 1,024 (2^{10}) levels per channel.

FRAME SIZE width and height of each frame.

CHARACTERISTICS OF DIGITAL VIDEO

BIT RATE is a measurement of the rate of information content from the digital video stream. The video size is proportional to the bit rate and the duration.

BITS PER PIXEL (BPP) is a measure of the efficiency of compression. A true-color video with no compression at all may have a BPP of 24 bits/pixel. Applying JPEG compression on every frame can reduce the BPP to 8 or 1 bits/pixel. Applying video compression algorithms like MPEG4 allows for fractional BPP values.

TABLE OF CONTENTS

- 1 COURSE LOGISTICS
- 2 MODULE 1
- 3 DIGITAL VIDEO
- 4 VIDEO ANALYTICS
- 5 APPLICATIONS OF VIDEO ANALYTICS
- 6 ANALOG VIDEO FORMATS
- 7 SPATIO TEMPORAL SAMPLING STRUCTURE

DIGITAL VIDEO PROCESSING

- Digital video processing is the study of algorithms for processing moving images that are represented in digital format.
- Data intensive
 - ▶ significant bandwidth
 - ▶ computational resources
 - ▶ storage resources

VIDEO ANALYTICS

- Video analytics (VA), is the capability of automatically analysing video to detect and determine temporal and spatial events.
- Also known as **Video content analysis or video content analytics (VCA)**, and **video analysis (VA)**
- VA applies computer vision to deal with the analysis of digital images and videos.

FUNCTIONALITIES IN VIDEO ANALYTICS

MOTION DETECTION motion is detected with regard to a fixed background scene.

DYNAMIC MASKING Blocking a part of the video signal based on the signal itself, for example because of privacy concerns.

EGOMOTION ESTIMATION is used to determine the location of a camera by analyzing its output signal.

SHAPE RECOGNITION recognize shapes in the input video.

OBJECT DETECTION determine the presence of a type of object or entity.

RECOGNITION Face recognition, Action recognition and Automatic Number Plate Recognition are used to recognize, and identify, persons, actions or cars, respectively.

FUNCTIONALITIES IN VIDEO ANALYTICS

TAMPER DETECTION is used to determine whether the camera or output signal is tampered with.

VIDEO TRACKING determine the location of objects in the video signal, with regard to an external reference grid.

OBJECT CO-SEGMENTATION Joint object discovery, classification and segmentation of targets in one or multiple related video sequences.

VIDEO SUMMARIZATION identify and extract from the original video content the most important frames (key-frames), and/or the most important video segments (key-shots), normally in a temporally ordered fashion.

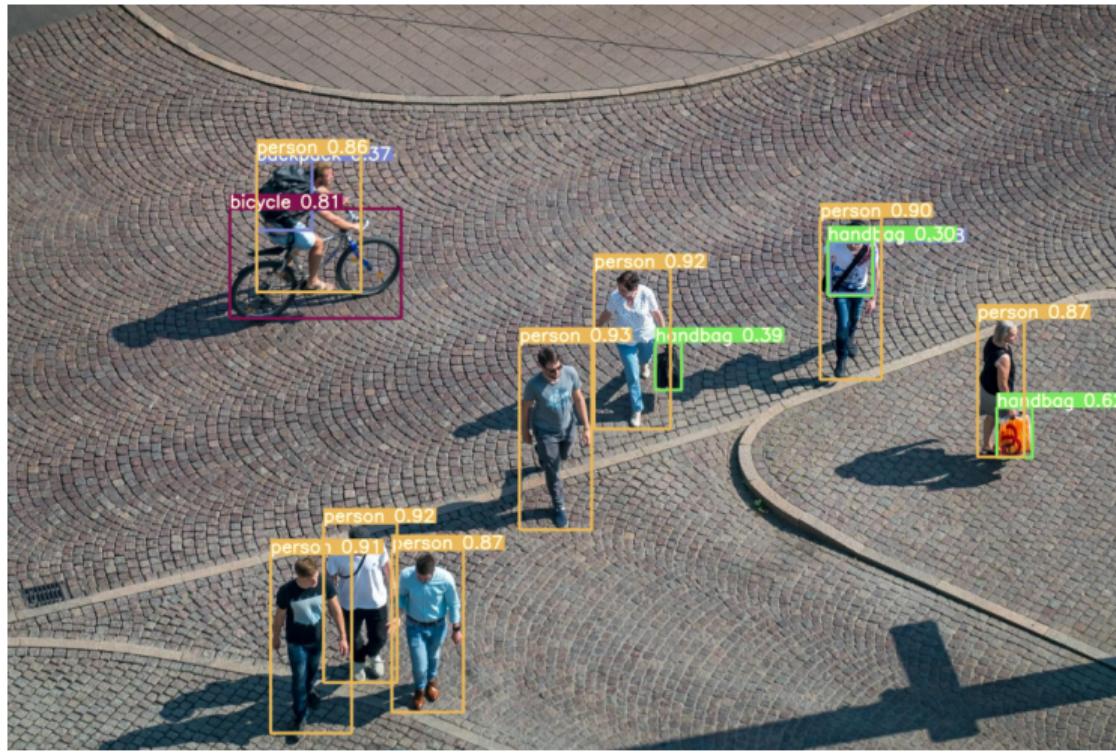
VIDEO SYNOPSIS automatically synthesize a short, informative summary of a video.

Detect Fake Videos

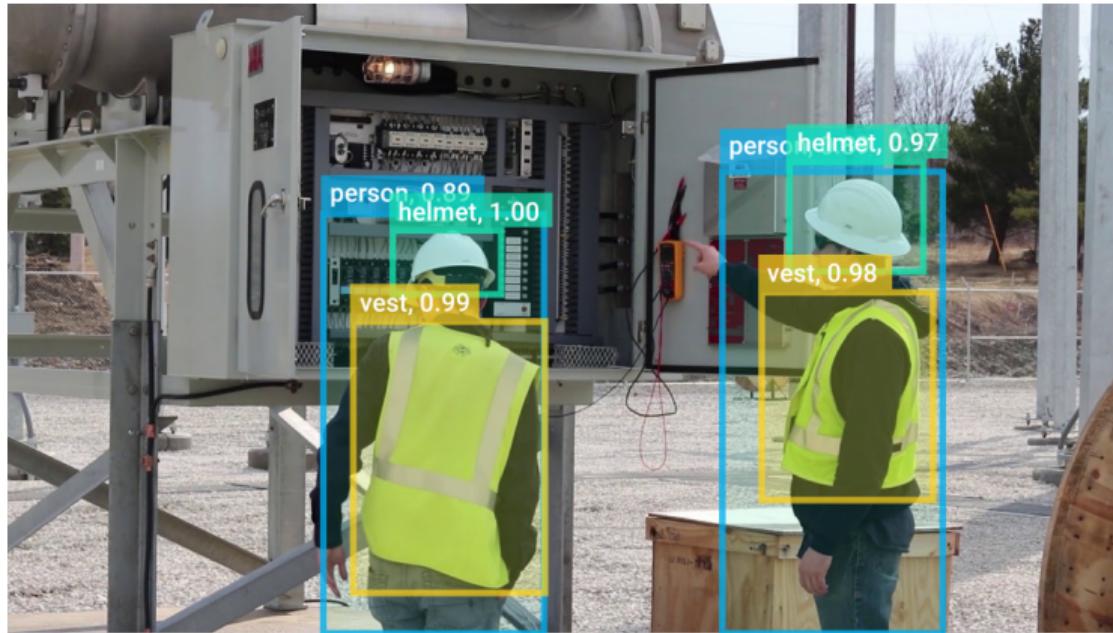
TABLE OF CONTENTS

- ① COURSE LOGISTICS
- ② MODULE 1
- ③ DIGITAL VIDEO
- ④ VIDEO ANALYTICS
- ⑤ APPLICATIONS OF VIDEO ANALYTICS
- ⑥ ANALOG VIDEO FORMATS
- ⑦ SPATIO TEMPORAL SAMPLING STRUCTURE

OBJECT DETECTION



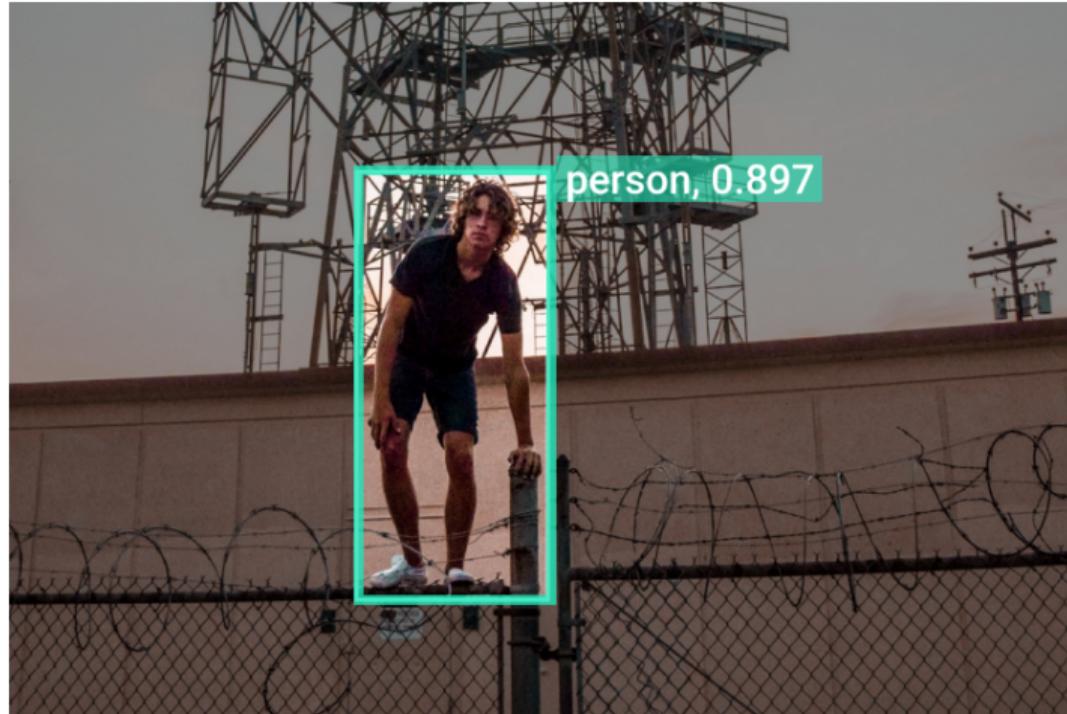
VEST / HELMET DETECTION IN SAFETY



PRODUCT DETECTION IN MANUFACTURING



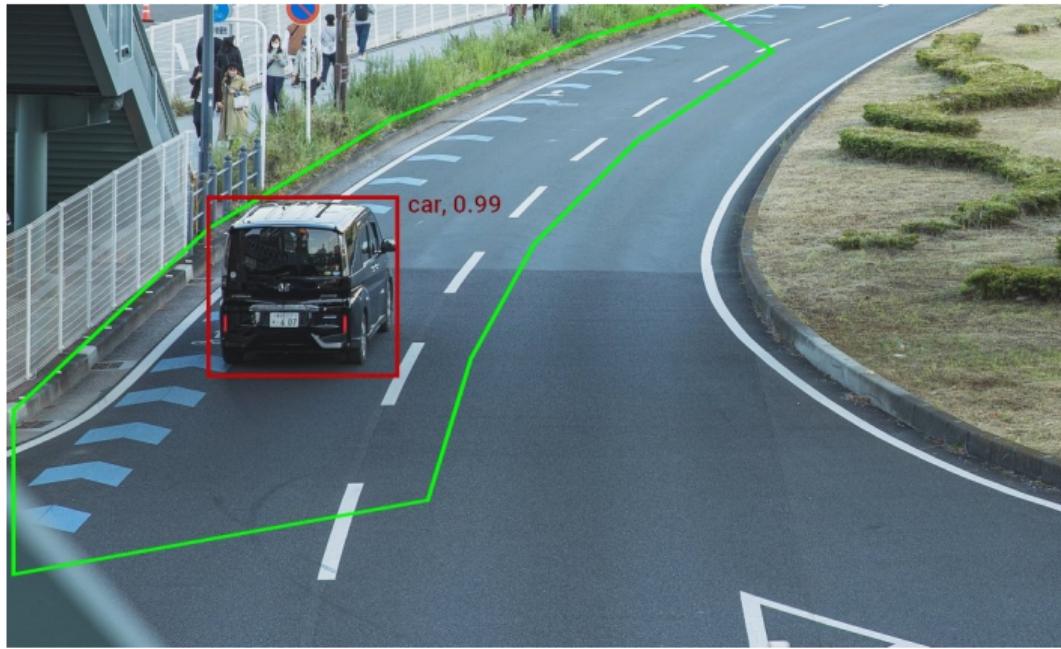
INTRUSION DETECTION FOR SECURITY



ABANDONED OBJECT DETECTION FOR SECURITY



STOPPED VEHICLE FOR SECURITY



QUEUE MANAGEMENT, FOOTFALL ANALYSIS, CUSTOMER ANALYTICS IN RETAIL



FALL DETECTION VEHICLE IN HEALTHCARE



EMOTION DETECTION



NUMBER PLATE DETECTION



TRAFFIC DETECTION

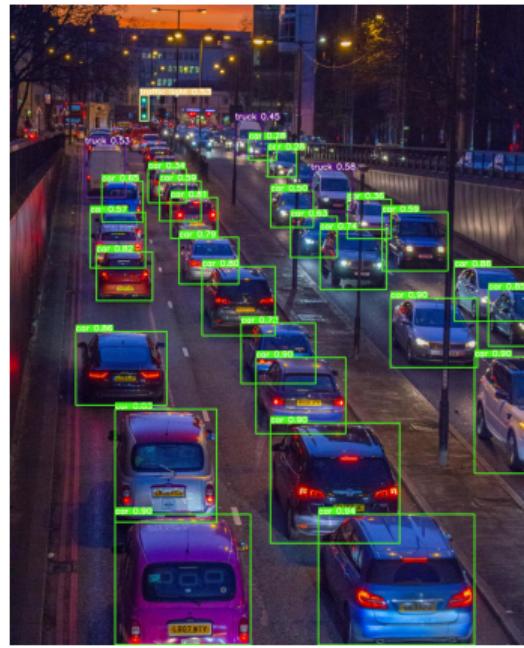


TABLE OF CONTENTS

- 1 COURSE LOGISTICS
- 2 MODULE 1
- 3 DIGITAL VIDEO
- 4 VIDEO ANALYTICS
- 5 APPLICATIONS OF VIDEO ANALYTICS
- 6 ANALOG VIDEO FORMATS
- 7 SPATIOTEMPORAL SAMPLING STRUCTURE

COMPOSITE ANALOG VIDEO

- Composite analog video combines brightness, color, synchronization into one signal.
- Formats are
 - ▶ National Television Standards Committee (NTSC) – 525 scan lines, 30 frames per second, 4:3 aspect ratio
 - ▶ Phase Alternate Line (PAL) – 625 scan lines, 25 frames per second, interlaced at 50 cycles per sec.
 - ▶ Sequential Color with Memory (SECAM) – 625 scan lines, interlaced at 50 cycles per sec.

COMPONENT VIDEO

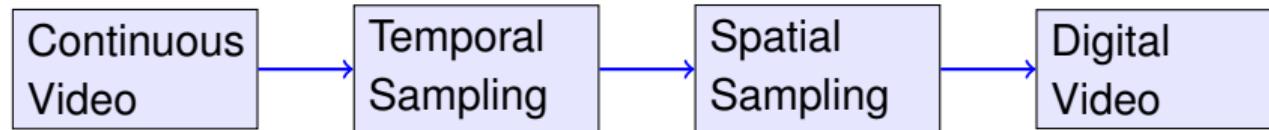
- Takes different components into separate channels.
- Formats are
 - ▶ S-VHS video – color and luminance information on two different channels.
 - ▶ RGB – separate channels for red, green and blue components.

TABLE OF CONTENTS

- 1 COURSE LOGISTICS
- 2 MODULE 1
- 3 DIGITAL VIDEO
- 4 VIDEO ANALYTICS
- 5 APPLICATIONS OF VIDEO ANALYTICS
- 6 ANALOG VIDEO FORMATS
- 7 SPATIOTEMPORAL SAMPLING STRUCTURE

DIGITAL VIDEO

- The picture information is digitized both spatially and temporally and the resultant pixel intensities are quantized.



DIGITAL VIDEO

- A continuous time-varying image $f_c(x, y, t)$ is a scalar real-valued function of two spatial dimensions x and y and time t .
- Let $pw \times ph$, where pw is the picture width and ph is the picture height.
- The ratio pw/ph is called the **aspect ratio**.
- The most common aspect ratio is 4/3 for standard TV and 16/9 for HDTV.

VIDEO SAMPLING AND QUANTIZATION

- **Video quantization** is essentially the same as image quantization.
- **Video sampling** involves taking samples along time dimension.
- Time proceeds from the past toward the future, with an origin that exists only in the **current moment**.

IMAGE SAMPLING AND QUANTIZATION

- The output of cameras / sensors is a continuous voltage waveform whose amplitude and spatial behavior are related to the physical phenomenon being sensed.
- Convert the continuous sensed data into digital image.
- A continuous image f may be continuous with respect to the x - and y -coordinates, and also in amplitude / intensity.
- To convert it to digital form, sample the function in both coordinates and in amplitude.
- Involves two processes
 - ▶ sampling – **Digitizing the coordinate values is called sampling.**
 - ▶ quantization – **Digitizing the amplitude values is called quantization.**

IMAGE SAMPLING AND QUANTIZATION

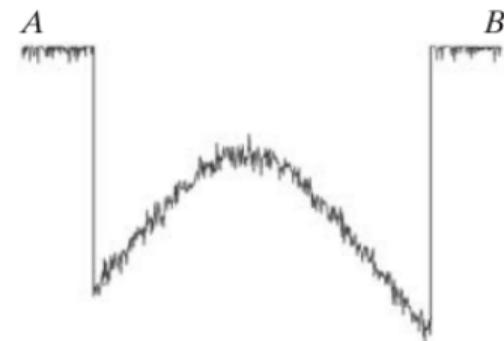
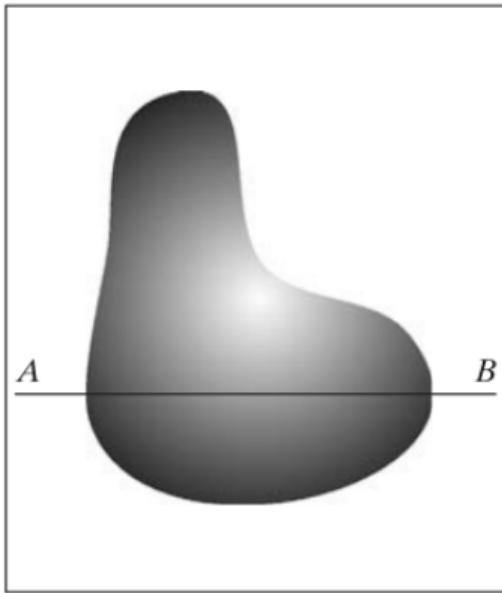


FIGURE: (a) Continuous image. (b) Plot of amplitude (intensity level) values of the continuous image along the line segment AB. The random variations are due to image noise.

IMAGE SAMPLING AND QUANTIZATION

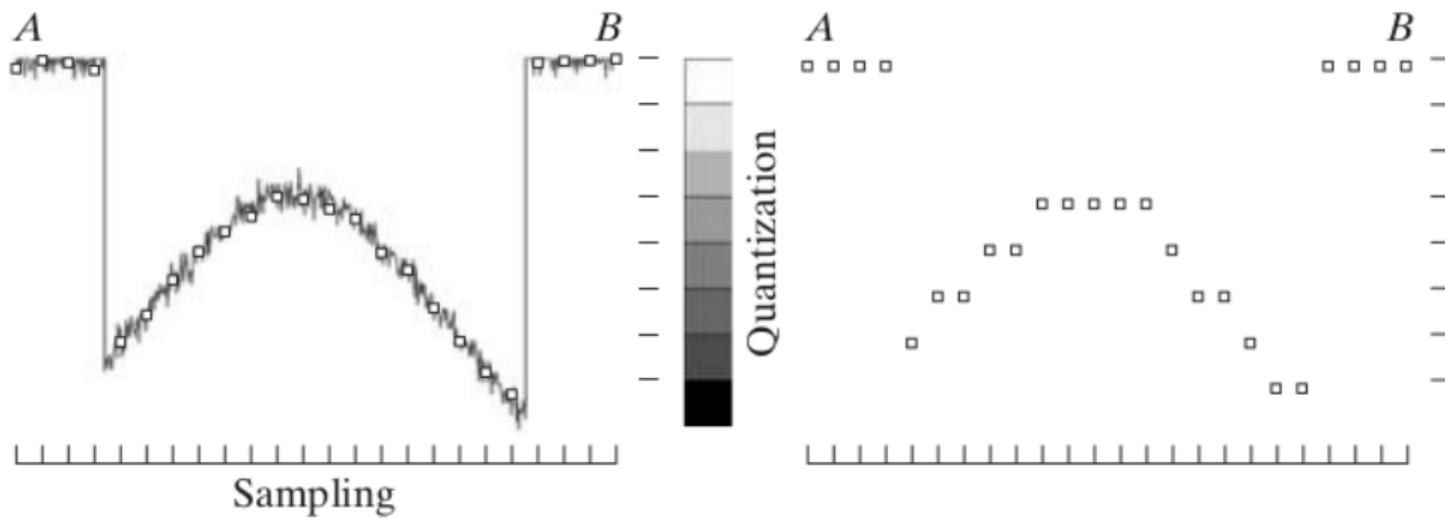


FIGURE: (a) Take equally spaced samples along line AB. The spatial location of each sample is indicated by a Vertical tick mark in the bottom part of the left side. The samples are shown as small white squares superimposed on the function.
(b) The right side of shows the intensity scale divided into eight discrete intervals, ranging from black to white. The vertical tick marks indicate the specific value assigned to each of the eight intensity intervals. The continuous intensity levels are quantized by assigning one of the eight values to each sample. The assignment is made depending on the vertical proximity of a sample to a vertical tick mark.

IMAGE SAMPLING AND QUANTIZATION

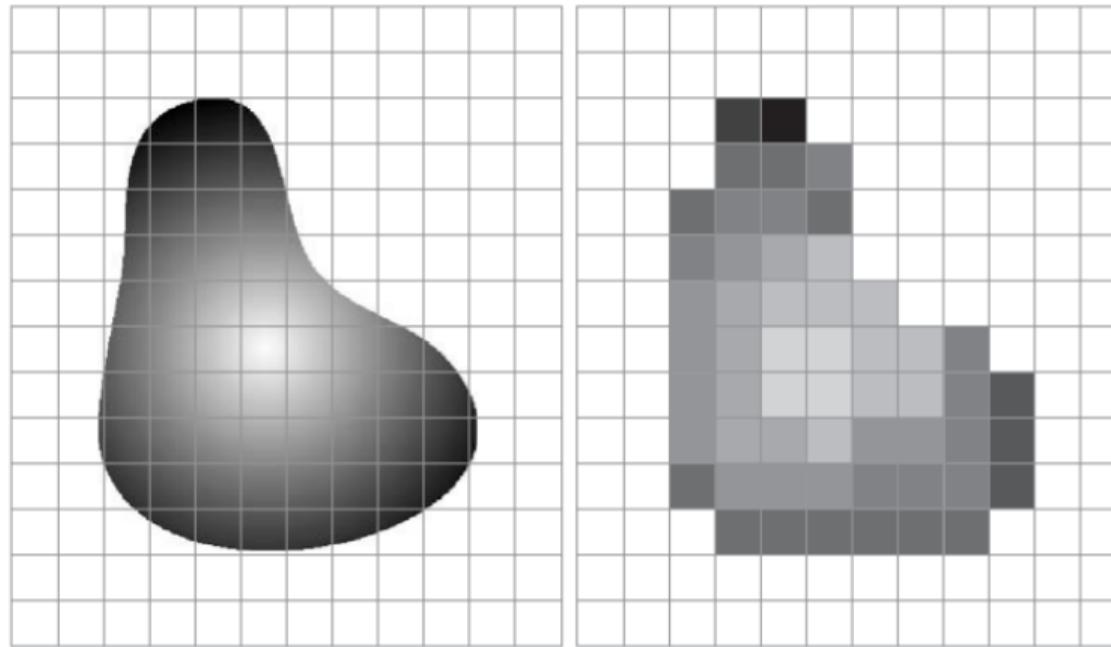


FIGURE: (a) Continuous image projected onto a sensor array. (b) Result of image sampling and quantization.

VIDEO SAMPLING

- Analog video systems, such as television and monitors, represent video as a one-dimensional electrical signal $V(t)$.
- Prior to display, a one-dimensional signal is obtained by sampling $I(x, y, t)$ along the vertical (y) space direction and along the time (t) direction.
- This is called **scanning** and the result is a series of time samples, which are complete pictures or **frames**, each of which is composed as space samples, or **scan lines**.
- **Refresh rate** is the frame rate at which information is displayed on a monitor.
 - ▶ Frame rate should be high enough, otherwise the displayed video will appear to **flicker**.
 - ▶ The human eye detects flicker if the refresh rate is less than about 50 frames/s.

VIDEO SCANNING

Two types of video scanning

- Progressive
 - ▶ A progressive scan traces a complete frame, line-by-line from top-to-bottom, at a scan rate of Δt s/frame.
 - ▶ High-resolution computer monitors are a good example, with a scan rate of $\Delta t = 1/72s$.
- Interlaced
 - ▶ In $P : 1$ interlacing, every P^{th} line is refreshed at each frame refresh. The subframes in interlaced video are called **fields**, and hence P fields constitute a frame.
 - ▶ The standard television systems uses $2 : 1$ interlacing. The two fields are usually referred to as the top and bottom fields. The flicker is effectively eliminated provided that the field refresh rate is above the visual limit of about 50 Hertz (Hz).
 - ▶ If an interlaced video has a frame rate of 30 frames per second, the field rate is

VIDEO SCANNING

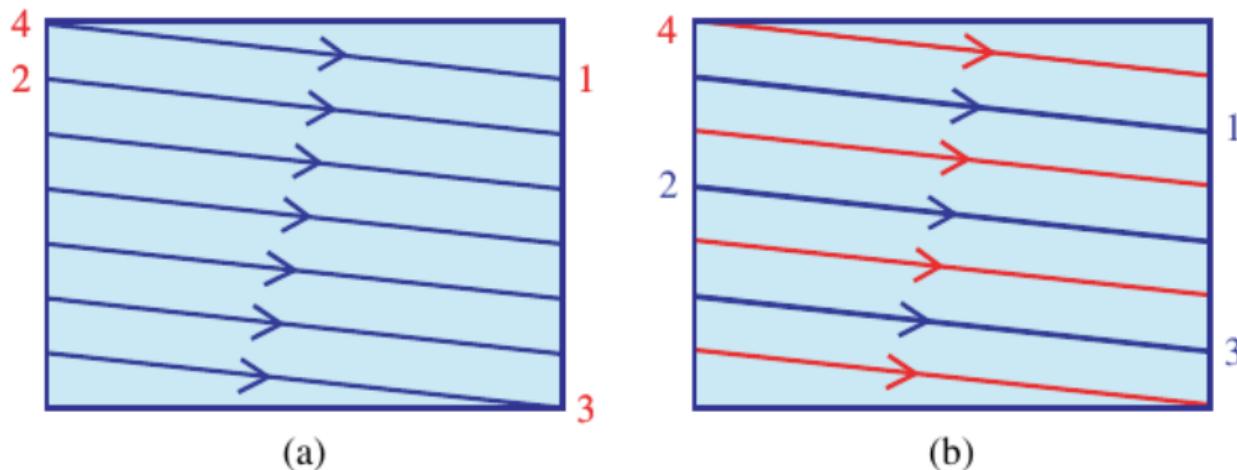


FIGURE: (a) Progressive video scanning. At the end of a scan (1), the electron gun spot snaps back to (2). A blank signal is sent in the interim. After reaching the end of a frame (3), the spot snaps back to (4). A synchronization pulse then signals the start of another frame. (b) Interlaced video scanning. Red and blue fields are alternately scanned left-to-right and top-to-bottom. At the end of scan (1), the spot snaps to (2). At the end of the blue field (3), the spot snaps to (4) (new field).

VIDEO SCANNING

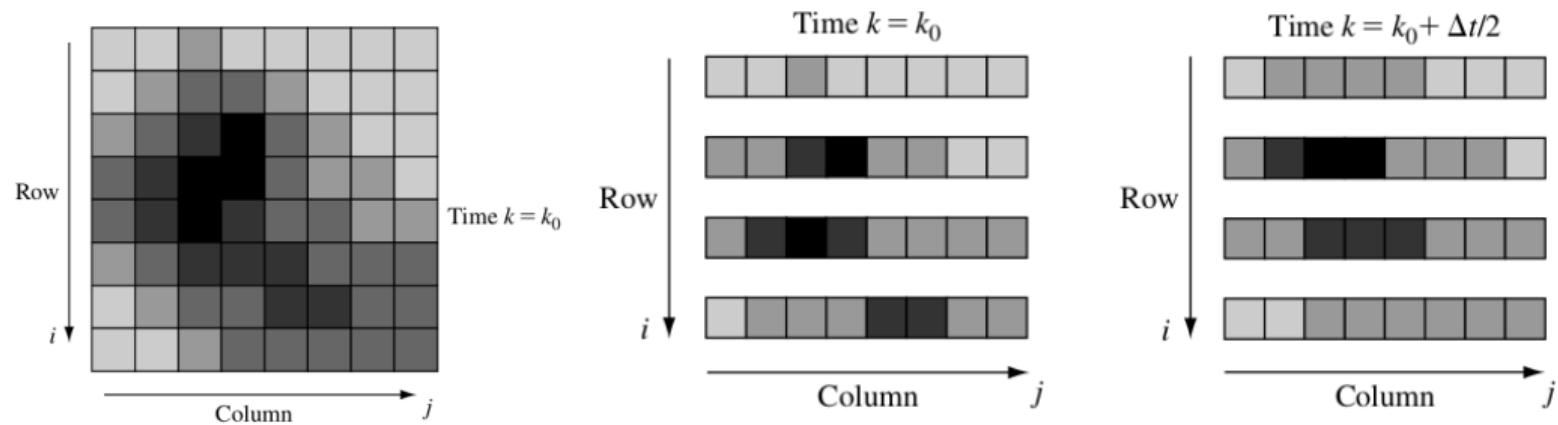


FIGURE: (a) A single frame from a sampled progressive video sequence. (b) A single frame (two fields) from a sampled 2:1 interlaced video sequence.

VIDEO SAMPLING

- The image f_c can be sampled in one, two, or three dimensions.
- When image is sampled in at least the temporal dimension, it produces an **image sequence**. An example of an image sampled only in the temporal dimension is motion picture film.
- Analog video is typically sampled in the vertical and temporal dimensions using one of the scanning structures. Example is monitor and television.
- Digital video is sampled in all three dimensions. The subset of \mathcal{R}^3 on which the sampled image is defined is called the sampling structure Ψ , is contained in \mathcal{W}_T . The mathematical structure used in describing sampling of time-varying images is the **lattice**.

VIDEO COMPRESSION PICTURE / FRAME TYPES

- A **video frame** is compressed using different algorithms, centered around amount of data compression.
- The different algorithms for video frames are called **picture types or frame types**.

FRAME TYPES

Three major types of frame types

I-FRAMES (**INTRA-CODED PICTURE**) is a complete image, like a JPG or BMP image file. A frame used as a reference for predicting other frames is called a reference frame or key-frame.

P-FRAMES (**PREDICTED PICTURE**) holds only the changes in the image from the previous frame. P-frames are also known as delta-frames. and are more compressible than I-frames.

B-FRAMES (**BIDIRECTIONAL PREDICTED PICTURE**) saves even more space by using differences between the current frame and both the preceding and following frames to specify its content.

FRAME TYPES

Characteristic	I-frame	P-frame	B-frame
Image	Complete image	use data from previous frames to decompress	use both previous and following frames for data reference
Compression	least compressible but don't require other video frames to decode.	more compressible than I-frame	highest amount of data compression

FRAME TYPES

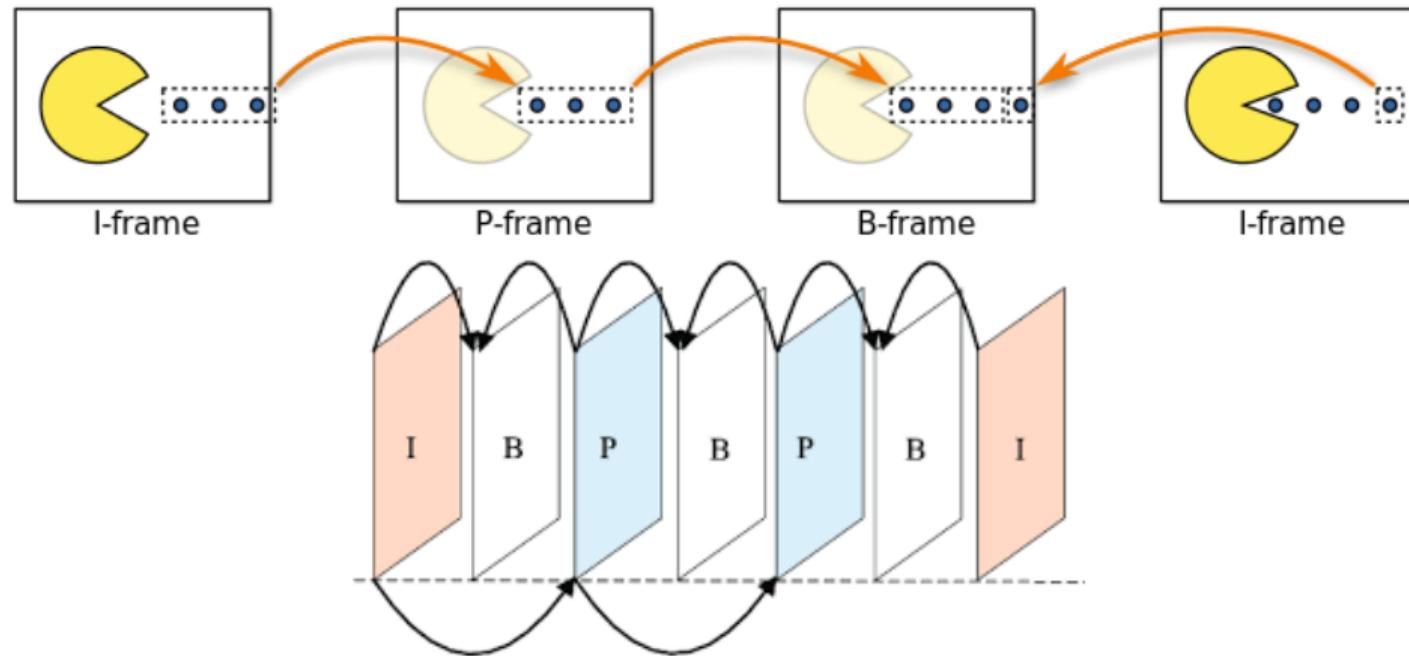


FIGURE: I-frame, P-frame, B-frame

FRAME TYPES – EXAMPLE

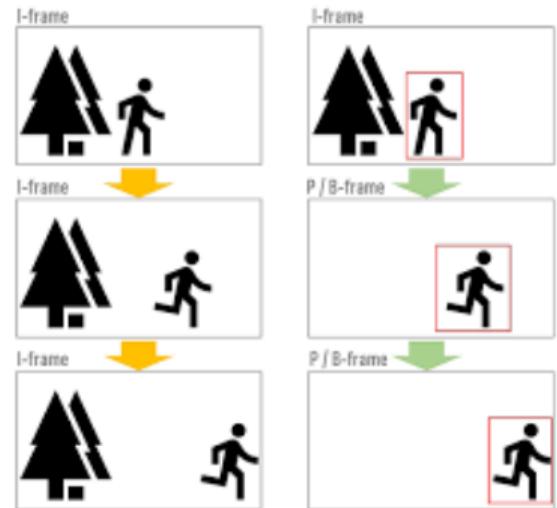


FIGURE: An Example of Frame types

PICTURE, FRAME, FIELD

PICTURE is a more general notion. A picture can be either a frame or a field.

FRAME is a complete image.

FIELD is the set of odd-numbered or even-numbered scan lines composing a partial image.

For example, an HD 1080 picture has 1080 lines (rows) of pixels. An odd field consists of pixel information for lines 1, 3, 5...1079. An even field has pixel information for lines 2, 4, 6...1080. When video is sent in interlaced-scan format, each frame is sent in two fields, the field of odd-numbered lines followed by the field of even-numbered lines.

FRAME, FIELD

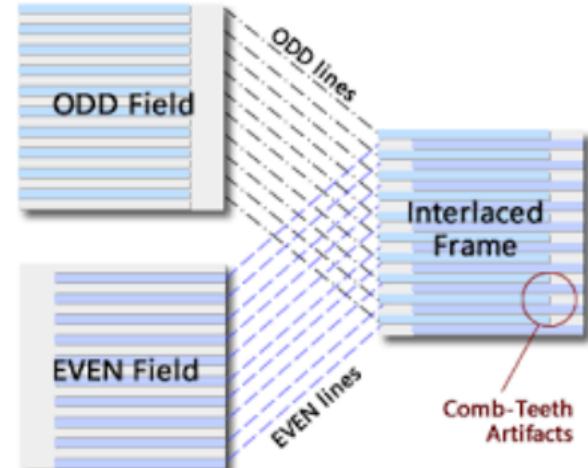
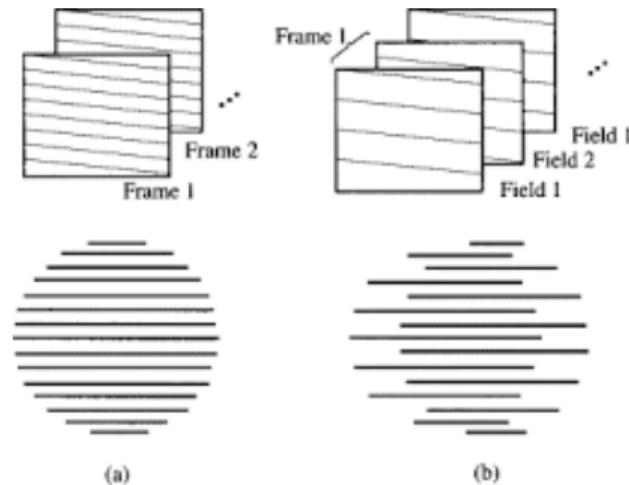


FIGURE: Frames and Fields

FRAME, FIELD – EXAMPLE

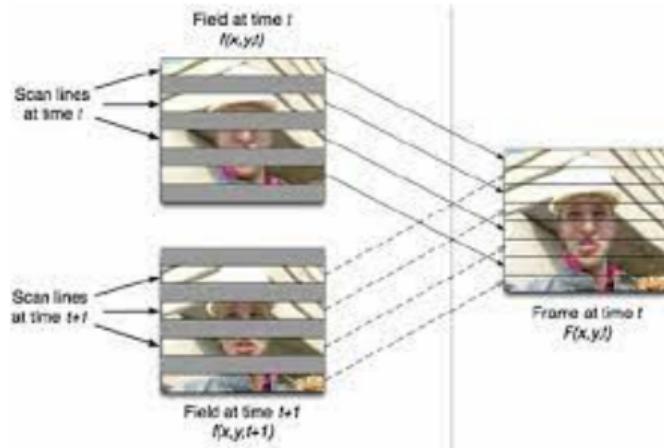


FIGURE: An Example of Frame and Field decomposition

FURTHER READING / VIEWING

- ① Digital Video Standards from R1 Ch1.2.2
- ② <https://www.toptal.com/machine-learning/machine-learning-video-analysis>
- ③ <https://cloud.google.com/video-intelligence>
- ④ https://youtu.be/RbR_yqg85FM
- ⑤ <https://youtu.be/1Hog27WNTRY>
- ⑥ <https://tryolabs.com/assets/guides/video-analytics-guide/people-detection-ebddd6b47f.gif>
- ⑦ https://tryolabs.com/assets/guides/video-analytics-guide/possession_compressed_1-6ba1f68fb2.mp4
- ⑧ <https://blog.video.ibm.com/streaming-video-tips/keyframes-interframe-video-compression/>

REFERENCES

- ① Bovik, Alan C. The essential guide to video processing. (T1) Ch 1
- ② Tekalp, A. Murat. Digital video processing. (R1) Ch 1.2
- ③ Rafael C. Gonzalez and Richard E. Woods, Digital Image Processing Ch 2.4
- ④ https://en.wikipedia.org/wiki/Video_content_analysis
- ⑤ https://en.wikipedia.org/wiki/Automatic_summarization
- ⑥ <https://en.wikipedia.org/wiki/VideoSynopsis>
- ⑦ https://en.wikipedia.org/wiki/Video_browsing
- ⑧ https://en.wikipedia.org/wiki/Video_compression_picture_types
- ⑨ <https://viso.ai/computer-vision/video-analytics-ultimate-overview/>

Thank You!



VIDEO ANALYTICS MODULE # 2 : MOTION DETECTION AND ESTIMATION

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

DL Team, BITS Pilani

The instructor is gratefully acknowledging
the authors who made their course
materials freely available online.

This deck is prepared by Seetha Parameswaran.

TABLE OF CONTENTS

- 1 MODULE 2
- 2 INTRODUCTION TO MOTION
- 3 MATH PRELIMINARIES
- 4 MOTION DETECTION
- 5 MOTION ESTIMATION

MODULE TOPICS....

- MRF and MAP
- Motion detection
- Motion estimation
- Optical Flow Motion estimation
- MAP estimation for Dense motion
- Application

TABLE OF CONTENTS

① MODULE 2

② INTRODUCTION TO MOTION

③ MATH PRELIMINARIES

④ MOTION DETECTION

⑤ MOTION ESTIMATION

VIDEO AND MOTION

- Video captures motion.
 - A single image provides snapshot of a scene.
 - A sequence of images records scene's dynamics.
- The recorded motion is a very strong cue for human vision.
 - Easy to recognize objects as soon as they move.

MOTION

- Motion is important for video processing and compression for two reasons.
 - ① Motion carries information about spatio-temporal relationships between objects in the field of view of a camera.
 - ★ used in applications such as traffic monitoring or security surveillance.
 - ② Image properties, such as intensity or color, have a very high correlation in the direction of motion. They do not change significantly when tracked over time
 - ★ used for the removal of temporal redundancy in video coding.
- Two-dimensional (2D) motion of intensity patterns in the image plane is referred to as **apparent motion**.

MOTION RELATED TASKS

MOTION DETECTION identify image points as moving or stationary.

MOTION ESTIMATION measure how image points move.

MOTION SEGMENTATION identify groups of image points moving similarly.

TABLE OF CONTENTS

1 MODULE 2

2 INTRODUCTION TO MOTION

3 MATH PRELIMINARIES

4 MOTION DETECTION

5 MOTION ESTIMATION

IMAGE SEQUENCE

- Intensity of image sequence

$$\mathcal{I} : \Omega \times \mathcal{T} \rightarrow R^+$$

- Ω - spatial domain
- \mathcal{T} - temporal domain
- $\mathbf{x} = (x_1, x_2)^T \in \Omega$ - Spatial position of a point in the image sequence
- $t \in \mathcal{T}$ - Temporal position of a point in the image sequence

MOTION

- $\nu = (\nu_1, \nu_2)^T$ - Velocity vector to represent motion in continuous images.
- ν_t - Dense velocity field or motion field, that is, the set of all velocity vectors within the image, at time t .
- \mathbf{b}_t - small number of motion parameters. Reduce computational complexity.

$$\nu_t \xrightarrow{\text{transformation}} \mathbf{b}_t$$

- \mathbf{d} - Displacement / velocity vector to represent motion in discrete images

CONTINUOUS VS DISCRETE REPRESENTATION

Representation	Continuous	Discrete
Time	t	t_k
Position	(\mathbf{x}, t)	$((n), t_k) = ((n_1, n_2)^T, t_k)$
Image	$I(\mathbf{x}, t)$	$I[\mathbf{n}, t]$
Motion	$\mathbf{v} = (\nu_1, \nu_2)^T$	

BINARY HYPOTHESIS TESTING

- Let y be an observation.
- Let Y be the associated random variable.
- Two hypotheses
 - ▶ H_0 - probability distributions $P(Y = y|H_0)$
 - ▶ H_1 - probability distributions $P(Y = y|H_1)$
- Goal: **Decide from which of the two distributions a given y is selected.**

BINARY HYPOTHESIS TESTING

- 4 possibilities for true hypothesis /decision
 - ▶ $H_0 \mid H_0$ - correct choice
 - ▶ $H_1 \mid H_1$ - correct choice
 - ▶ $H_0 \mid H_1$ - error
 - ▶ $H_1 \mid H_0$ - error
- To make a decision, a decision criterion is needed that attaches some relative importance to the four possibilities. i.e assign cost to each decision.

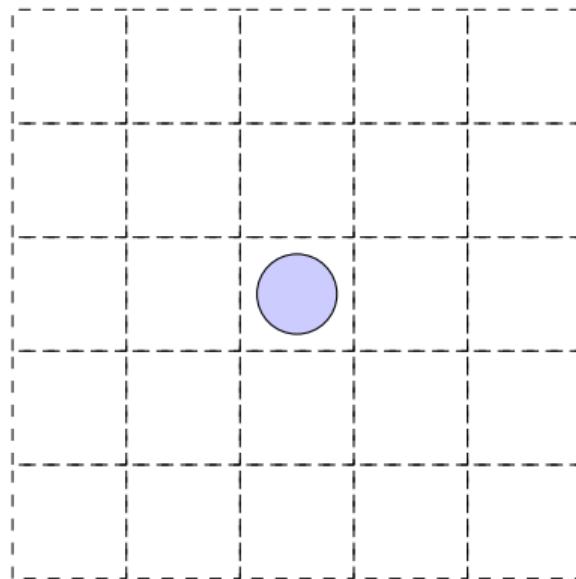
BAYES CRITERION

- Two a priori probabilities
 - ▶ π_0 for H_0
 - ▶ $\pi_1 = 1 - \pi_0$ for H_1
- Design a decision rule so that on average the cost associated with making a decision based on y is minimal.
- Optimal decision can be made according to the following rule

$$\frac{P_1}{P_0} = \underbrace{\frac{P(y = y | H_1)}{P(Y = y | H_0)}}_{\text{Likelihood ratio}} \gtrless \underbrace{\vartheta}_{\text{constant}} \quad \underbrace{\frac{\pi_0}{\pi_1}}_{\text{prior probability}}$$

SAMPLING GRID

- Let Λ be a sampling grid in R^N .
- Let $n \in \Lambda$ be any point the grid.

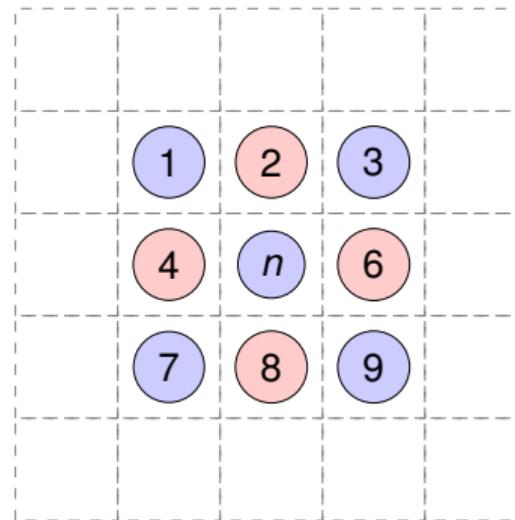


SAMPLING GRID AND NEIGHBORHOOD

- Let $n \in \Lambda$ be any point the grid.
- Let $\eta(n)$ be a neighborhood of $n \in \Lambda$.
- The first-order neighborhood consists of immediate top, bottom, left, and right neighbors of n .
- Let \mathcal{N} be a neighborhood system, a collection of neighborhoods of all $n \in \Lambda$.

$$n \notin \eta(n)$$

$$n \in \eta(l) \Leftrightarrow l \in \eta(n)$$



RANDOM FIELD

- A random field \mathcal{Y} over Λ is a multidimensional random process where each site $n \in \Lambda$ is assigned a random variable.
- A random field \mathcal{Y} with the following properties:

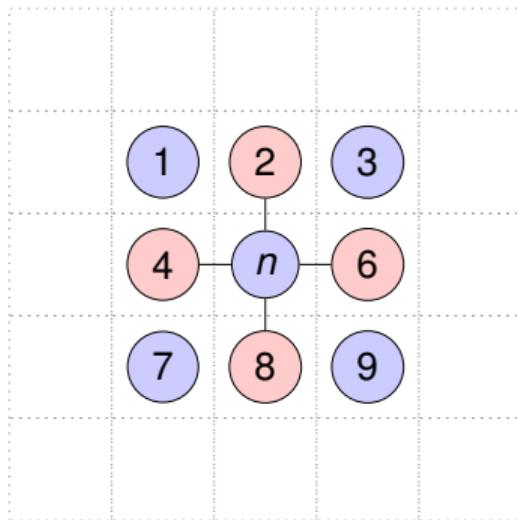
$$P(\mathcal{Y} = \nu) > 0 \quad \forall \nu \in \Gamma$$

$$P(\mathcal{Y}_n = \nu_n \mid \mathcal{Y}_I = \nu_I, \forall I \neq n) = P(\mathcal{Y}_n = \nu_n \mid \mathcal{Y}_I = \nu_I, \forall I \in \eta(n)) \quad \forall n \in \Lambda, \forall \nu \in \Gamma$$

P is a probability measure, and is called a Markov random field with state space Γ .

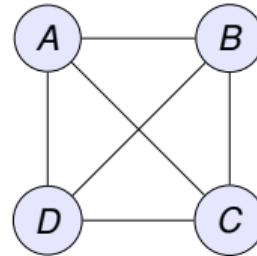
CLIQUE

- A clique c defined over Λ with respect to \mathcal{N} is a subset of Λ such that either c consists of a single site or every pair of sites in c are neighbors, that is, belong to η .
- The set of all cliques is denoted by C .
- A two-element spatial clique $\{n, l\}$ are two immediate horizontal, vertical or diagonal neighbors.



CLIQUE IN GENERAL

- Clique is a subset of vertices of an undirected graph such that every two distinct vertices in the clique are adjacent.
- A **clique** is a subset of nodes in which every node is connected to every other node.
- A **maximal clique** is a clique which cannot be extended by the addition of another node.



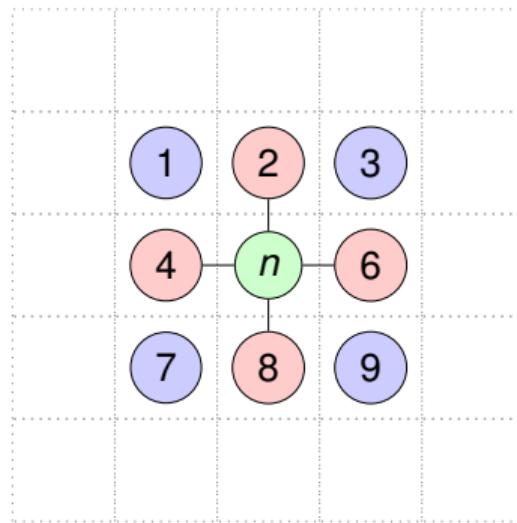
MARKOV RANDOM FIELD

- A Markov Random Field (MRF) is a graphical model of a joint probability distribution.
- Edges encode conditional independence.
- Let X_S be the set of random variables associated with the set of nodes S .
- Rule: Given disjoint subsets of nodes A , B and C , X_A is conditionally independent of X_B given X_C if there is no path from any node in A to any node in B that doesn't pass through a node of C . relationships.
- Markov property tells us that the joint distribution of X is determined entirely by the **local conditional distributions** $P(X_n \mid X_{\eta(n)})$

MARKOV RANDOM FIELD

Given the pink nodes, the green node is conditionally independent of all other nodes.

$$n \perp \{1, 3, 7, 9\} \mid \{2, 4, 6, 8\}$$



GIBBS DISTRIBUTION

- A Gibbs distribution on grid Λ and neighborhood \mathcal{N} takes the form:

$$P(\mathcal{Y} = \nu) = \frac{1}{Z} \exp\left(\frac{-1}{T} U(\nu)\right)$$

Z - Partition Function

T - temperature and is often taken to be 1.

U - energy function or potential function

$$U(\nu) = \sum_{c \in C} V(\nu, c)$$

For two element clique $\{n, l\}$ then $U(n, l) = V(\nu[n], \nu[l])$

MRF EQUIVALENCE TO GIBBS DISTRIBUTION

- The equivalence between Markov random fields and Gibbs distributions is provided through the important **Hammersley-Clifford theorem**.
- Theorem states that \mathcal{Y} is a MRF on Λ with respect to \mathcal{N} if and only if its probability distribution is a Gibbs distribution with respect to Λ and \mathcal{N} .

MAP ESTIMATION

- Let Y be a random field of observations
- Let \mathcal{Y} be a random field modeling the quantity we want to estimate based on Y
- .
- Let y, ν be their respective realizations.
- y could be a difference between two images.
- ν could be a field of motion detection labels.
- To compute ν based on y (MAP) estimation is

$$\hat{\nu} = \arg \max_{\nu} \underbrace{P(\mathcal{Y} = \nu \mid y)}_{\text{maximum of the posterior probability}} = \arg \max_{\nu} \underbrace{P(\mathcal{Y} = y \mid \nu) \cdot P(\mathcal{Y} = \nu)}_{\text{maximum likelihood}}$$

ML ESTIMATION

- To compute ν based on y (MAP) estimation is

$$\hat{\nu} = \arg \max_{\nu} \underbrace{P(\mathcal{Y} = \nu \mid y)}_{\text{maximum of the posterior probability}} = \arg \max_{\nu} \underbrace{P(\mathcal{Y} = y \mid \nu) \cdot P(\mathcal{Y} = \nu)}_{\text{maximum likelihood}}$$

- If $P(\mathcal{Y} = \nu)$ is the same for all realizations ν , then only the likelihood $P(\mathcal{Y} = y \mid \nu)$ is maximized, resulting in the maximum likelihood (ML) estimation.

TABLE OF CONTENTS

- 1 MODULE 2
- 2 INTRODUCTION TO MOTION
- 3 MATH PRELIMINARIES
- 4 MOTION DETECTION
- 5 MOTION ESTIMATION

GOAL OF MOTION DETECTION

- Identify which image points, or, regions of the image, have moved.
- Motion detection applies to images acquired with a **static camera**.
- Motion of image points is not perceived directly but rather through intensity changes.
- Such intensity changes over time may be also induced by camera noise or illumination variations.

HYPOTHESIS TESTING

- Simplest motion detection algorithms
- Let H_S and H_M be two hypotheses.
- H_S declaring an image point at n as stationary (S). State S means 0.
- H_M says an image point at n is moving (M). State M means 1.
- Let q be noise. Noise is assumed as zero-mean Gaussian with variance σ^2 in stationary areas and uniformly distributed in range $[-L, L]$ in moving areas.
- Assume $I_k[n] = I_{k-1}[n] + q$.
- P_S is assumed Gaussian, while P_M is assumed uniform.
- The motivation is that in stationary areas only camera noise will distinguish same-position pixels, whereas in moving areas this difference is attributed to motion and therefore unpredictable.

HYPOTHESIS TESTING WITH FIXED THRESHOLD

- Let an observation, upon which we intend to select one of the two hypotheses be

$$\rho_k[n] = I_k[n] - I_{k-1}[n]$$

- Hypothesis test is:

$$\rho_k^2[n] > \theta \quad \forall n \in M$$

$$\rho_k^2[n] < \theta \quad \forall n \in S$$

$$\theta = 2\sigma^2 \ln \left(\vartheta \cdot 2L \cdot P_S / (\sqrt{2\pi\sigma^2} \cdot P_M) \right)$$

- Not robust to noise in the image
- For small θ "noisy" detection masks result
- For large θ only object boundaries and its most textured parts are detected.

HYPOTHESIS TESTING, FIXED THRESHOLD, AVERAGING

- Averaging the observations over an N -point spatial window W_n centered at n

$$\frac{1}{N} \sum_{m \in W_n} \rho_k^2[m] > \theta \quad \forall n \in M$$

- Used to attenuate the impact of noise

MOTION DETECTION BASED ON FRAME DIFFERENCES

- Motion detection based on frame differences (last 3 slides) does not perform well for large, untextured objects (e.g., a large, uniformly colored truck).
- Only pixels n where $|I_k[n] - I_{k-1}[n]|$ is sufficiently large can be reliably detected.
- Such pixels concentrate in narrow areas close to moving boundaries where object intensity is distinct from the background in the previous frame.

MOTION DETECTION BY COMPARING BACKGROUND INTENSITY

- Comparing the current intensity $I_k[n]$ to background intensity $B_k[n]$ instead of the previous frame $I_{k-1}[n]$.

$$\rho_k[n] = I_k[n] - B_k[n]$$

- Estimate $B_k[n]$ by means of temporal averaging or median filtering the intensity at each n .
- Median can suppress intensities associated with moving objects (large window) and is fast. It fails in the presence of parasitic motion, such as fluttering leaves or waves on water surface.

HYPOTHESIS TESTING WITH FIXED THRESHOLD

Non-parametric distributions

- At each location n of frame k , an estimate of the stationary (background) probability distribution is computed from K recent frames as

$$P_S(I_k[n]) = \frac{1}{K} \sum_{i=1}^K \kappa(I_k[n] - I_{k-i}[n])$$

κ is a zero-mean Gaussian with variance σ^2 considered as constant

- Hypothesis test:

$$P_S(I_k[n]) > \theta \quad \forall n \in S$$

$$P_S(I_k[n]) < \theta \quad \forall n \in M$$

$$\theta = \frac{P_M}{2L\vartheta P_S}$$

MAP MRF FORMULATION

- To find a MAP estimate of the random field E_k , maximize the posterior probability

$$P(E_k = e_k \mid \rho_k)$$

- Let $|I_k[n] - I_{k-1}[n]|$ be an observation modelled as

$$\rho_k[n] = \xi(e_k[n]) + q[n]$$

where q is zero-mean uncorrelated Gaussian noise with variance σ^2

$$\xi(e_k[n]) = \begin{cases} 0 & \text{if } e_k[n] = \mathcal{S} \\ \alpha & \text{if } e_k[n] = \mathcal{M} \end{cases}$$

α is average temporal intensity difference based on previous-time moving labels e_{k-1}

MAP MRF FORMULATION

- ξ attempts to closely model the observations since for a static image point it is zero, whereas for a moving point it tracks average temporal intensity mismatch.
- Likelihood $P(R_k = \rho_k | e_k)$
- Gibbs distribution for the a priori probability $\pi(E_k = e_k)$
- The overall energy function

$$U(\rho_k, e_{k-1}, e_k) = \frac{1}{2\sigma^2} \sum_n ((\rho_k[n] - \xi(e_k[n]))^2 + \sum_{[n,l] \in C} V_s(e_k[n], e_k[l]) + \sum_{[t_{k-1}, t_k]} V_t(e_{k-1}[n], e_k[n]))$$

MRF FORMULATION

- The first term measures how well each label at n explains the observation $\rho_k[n]$.
- The other terms measure how contiguous the labels are in the image plane (V_s) and in time (V_t).
- A simple MRF model supported on the second-order neighborhood with two-element cliques $c = [n, l]$

$$V(e_k[n], e_k[l]) = \begin{cases} 0 & \text{if } e_k[n] = e_k[l] \\ \beta & \text{if } e_k[n] \neq e_k[l] \end{cases}$$

TABLE OF CONTENTS

- 1 MODULE 2
- 2 INTRODUCTION TO MOTION
- 3 MATH PRELIMINARIES
- 4 MOTION DETECTION
- 5 MOTION ESTIMATION

CONCEPT OF MOTION

- Video Compression
 - ▶ Reduce the number of bits needed to represent a video sequence.
 - ▶ Estimated motion parameters should lead to the highest compression ratio possible.
- Video Processing
 - ▶ Methods that improve quality
 - ▶ Eg: motion-compensated noise reduction, motion-compensated interpolation, and motion-based video segmentation.
 - ▶ True motion of image points is estimated.
 - ▶ Eg: In motion-compensated temporal interpolation, the task is to compute new images located between existing images of a video sequence). The new images should be consistent with the existing ones, image points belonging to moving objects must be displaced according to the true motion as otherwise "jerky" motion of objects would result.

MOTION ESTIMATION ALGORITHM

Three important elements

① Motion Models

- ① Spatial Motion Model
- ② Temporal Motion Model
- ③ Region of Support
- ④ Observation Model

② Estimation criteria

- ① Pixel-Domain Criteria
- ② Regularization

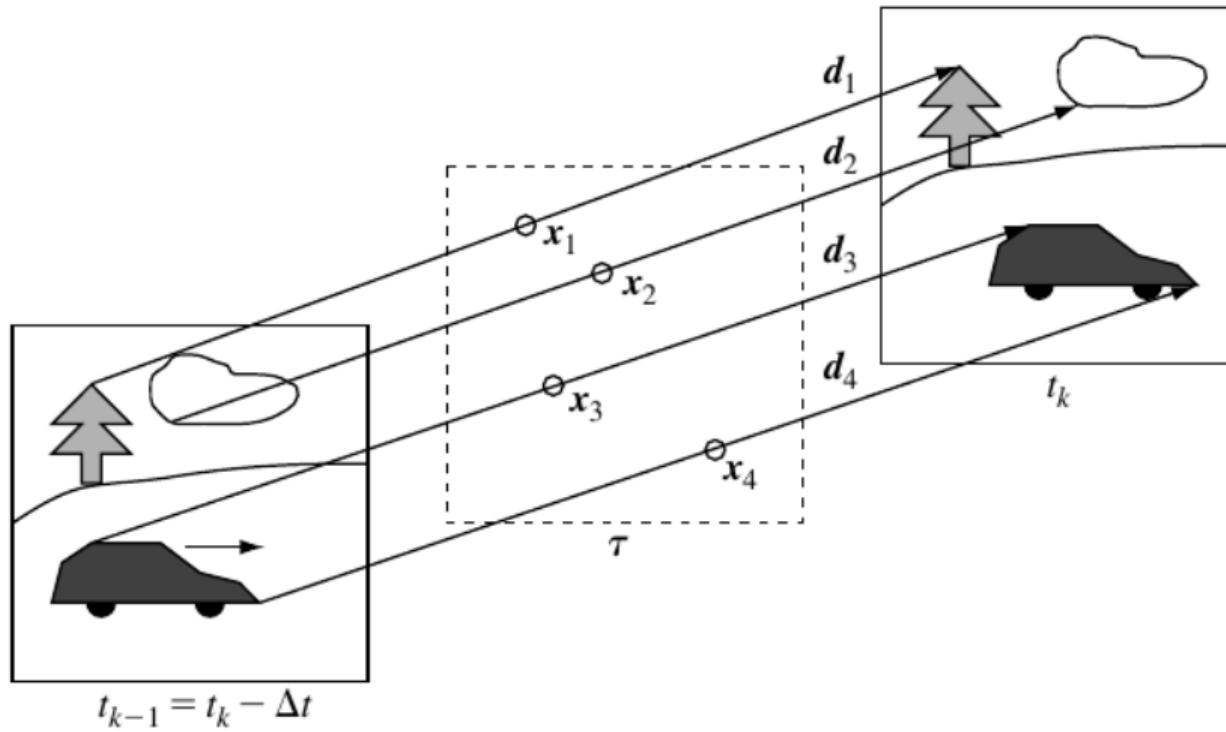
③ Search Strategies

- ① Matching
- ② Gradient-based techniques

MOTION MODEL

- A motion model specifies how to represent motion in an image sequence.
- A model relating motion parameters to image intensities is called an observation model.

MOTION MODEL



SPATIAL MOTION MODELS

- **The goal is to estimate the motion of image points.**
- Object-induced motion depends on the following:
 - ① image formation model, for example, perspective, orthographic projection
 - ② motion model of 3D object, for example, rigid-body with 3D translation and rotation, 3D affine motion
 - ③ surface model of 3D object, for example, planar, parabolic.

SPATIAL MOTION MODEL – TRANSLATIONAL MODEL

- Orthographic projection and arbitrary 3D surface undergoing 3D translation

Velocity vector $\nu(\mathbf{x}) = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$

Parameters $b = (b_1, b_2)^\top$

- Relatively simple and extensively used in practice.

SPATIAL MOTION MODEL – PARAMETRIC MODEL

- Orthographic projection and 3D affine motion of a planar surface

$$\text{Velocity vector } \nu(\mathbf{x}) = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} + \begin{pmatrix} b_3 & b_4 \\ b_5 & b_6 \end{pmatrix} \mathbf{x}$$
$$\text{Parameters } b = (b_1, \dots, b_6)^\top$$

- Simple and powerful
- A complex model applied to a small region of support may lead to an actual increase in the estimation error compared to a simpler model.

TEMPORAL MOTION MODELS

- The trajectories of individual image points drawn in the (x, y, t) space of an image sequence depend on object motion.
- Assume that the velocity $\nu_t(\mathbf{x})$ is constant between $t = t_{k-1}$ and $\tau (\tau > t)$
- $\mathbf{d}_{t,\tau}(\mathbf{x})$ is a displacement vector measured in the positive direction of time, from t to τ .
- The task is to find the two components of velocity or displacement at each \mathbf{x} .

TEMPORAL MOTION MODELS

- Linear Trajectory (two velocity (linear) variables)

$$\begin{aligned}\mathbf{x}(\tau) &= \mathbf{x}(t) + \nu_t(\mathbf{x}) \cdot (\tau - t) \\ &= \mathbf{x}(t) + \mathbf{d}_{t,\tau}(\mathbf{x})\end{aligned}$$

- Quadratic Trajectory (two velocity (linear) variables and two acceleration (quadratic) variables $\mathbf{a} = (a_1, a_2)^\top$)

$$\mathbf{x}(\tau) = \mathbf{x}(t) + \nu_t(\mathbf{x}) \cdot (\tau - t) + \frac{1}{2} \cdot \mathbf{a}_t(\mathbf{x}) \cdot (\tau - t)^2$$

REGION OF SUPPORT

- The set of points \mathbf{x} to which spatial and temporal motion models apply is called the region of support \mathcal{R} .
- For a given motion model, the smaller the region of support , the better the approximation of motion.
- Four types
 - ① $\mathcal{R} = \text{the whole image}$
 - ★ A single motion model applies to all image points.
 - ★ Most constrained model
 - ★ Very few parameters can approximate the motion of all image points.

REGION OF SUPPORT

② \mathcal{R} = one pixel

- ▶ This model applies to a single image point.
- ▶ Also called dense motion representation.
- ▶ Translational spatial model is used jointly with the linear temporal model.

Velocity vector $\nu(\mathbf{x}) = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$

Linear Trajectory $\mathbf{x}(\tau) = \mathbf{x}(t) + \nu_t(\mathbf{x}) \cdot (\tau - t)$

- ▶ Least constrained model
- ▶ A very large number of motion fields can be represented by all possible combinations of parameter values.
- ▶ High computational complexity

REGION OF SUPPORT

③ \mathcal{R} = rectangular block of pixels

- ▶ This motion model applies to a rectangular (or square) block of image points.
- ▶ Spatial translation model and temporal linear model of a square block of pixels is very powerful model and is used today in all digital video compression standards.
- ▶ Spatial translation and temporal quadratic motion is used in B frames of MPEG.

④ \mathcal{R} = irregularly-shaped region

- ▶ This model applies to all pixels in region of arbitrary shape.
- ▶ A square block divided into arbitrarily shaped parts, each with independent translational motion, is used in MPEG-4.

OBSERVATION MODELS

- Goal is to estimate motion based on intensity variations in time.
- Assume that objects do not change their appearance as they move. That is, image intensity $I = f(x, y, t)$ remains constant along motion trajectory s .

$$\frac{dI}{ds} = 0$$

$$\frac{\partial I}{\partial x} \nu_1 + \frac{\partial I}{\partial y} \nu_2 + \frac{\partial I}{\partial t} = 0 \quad \text{apply chain rule}$$

$$(\nabla I)^\top \nu + \frac{\partial I}{\partial t} = 0$$

- For I sampled in time, the constant-intensity assumption means that

$$I_{t_k}(x(t_k)) = I_{t_{k-1}}(x(t_{k-1}))$$

OBSERVATION MODELS

- Assume spatial sampling of intensities , linear trajectory, $t = t_{k-1}$ and $\tau = t_k$

$$\mathbf{x}(\tau) = \mathbf{x}(t) + \nu_t(\mathbf{x}) \cdot (\tau - t)$$

$$\frac{d\mathbf{I}}{ds} = I_k[n] - I_{k-1}[n - \mathbf{d}[n]] = 0, \quad \forall n \in \Lambda$$

- The above equation does not hold exactly due to noise, aliasing, illumination variations, etc., and a minimization of some function of the above equation is needed. When scene illumination changes, a constraint based on the spatial gradient's constancy in the direction of motion can be used.
- In areas of uniform intensity but substantial color detail, the inclusion of a color-based constraint could prove beneficial. A multicomponent (vector) function replaces \mathbf{I} .

MOTION ESTIMATION ALGORITHM

Three important elements

① Motion Models

- ① Spatial Motion Model
- ② Temporal Motion Model
- ③ Region of Support
- ④ Observation Model

② Estimation criteria

- ① Pixel-Domain Criteria
- ② Regularization

③ Search Strategies

- ① Matching
- ② Gradient-based techniques

ESTIMATION CRITERIA

- Motion models are incorporated into an estimation criterion that will be optimized.
- There is no unique criterion for motion estimation because its choice depends on the task at hand.
- In compression an average performance or prediction error of a motion estimator is the criteria.
- In motion-compensated interpolation the worst case performance (maximum interpolation error) is the criteria.
- The selection of a criterion may be guided by the processor capabilities on which the motion estimation will be implemented.

PIXEL-DOMAIN CRITERIA

- Motion-compensated prediction of $I_k[n]$ is given by

$$\tilde{I}_k[n] = I_{k-1}[n - d[n]]$$

- Then Error is given as

$$\epsilon_k[n] = I_k[n] - \tilde{I}_k[n] \quad \forall n \in \Lambda$$

- Discrete version of the constant-intensity assumption aim at the minimization of a function $\epsilon_k[n]$.
- Similarity between $I_k[n]$ and its prediction $\tilde{I}_k[n]$ can be also measured by the following cross-correlation function (which needs to be maximised):

$$C(d) = \sum_n I_k[n] \tilde{I}_{k-1}[n - d[n]]$$

PIXEL-DOMAIN CRITERIA

- Estimation criterion is then

$$\mathcal{E}(d) = \sum_{n \in R} \Phi(I_k[n] - \tilde{I}_k[n])$$

- Φ is a non-negative real-valued function.
- Quadratic function – a single large error ϵ (an outlier) over-contributes to \mathcal{E} and biases the estimate of \mathbf{d} .

$$\Phi(\epsilon) = \epsilon^2$$

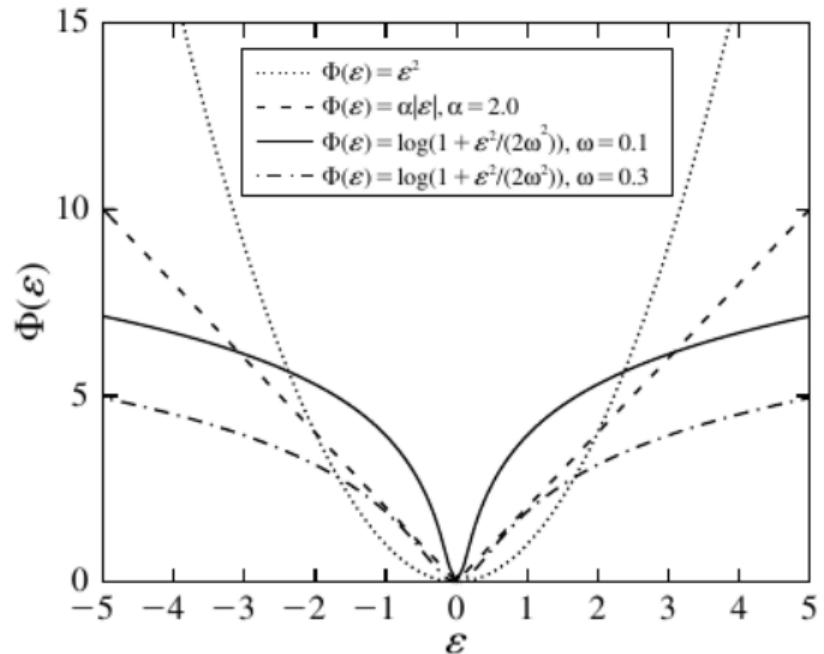
- Linear function – better choice, used in video encoders

$$\Phi(\epsilon) = \alpha(\epsilon)$$

- Lorentzian function – grows slower than $|x|$ for large errors.

$$\Phi(\epsilon) = \log(1 + \epsilon^2/2\omega^2)$$

PIXEL-DOMAIN ESTIMATION CRITERIA



REGULARIZATION

- Moving objects are close to being rigid.
- Motion field ν_t is locally smooth.
- Gradient is a good measure of local smoothness.
- Minimizing the following criterion : where D is the domain of the image. This formulation is often referred to as regularization.

$$E(\nu) = \int_D \left(\nabla^\top I(x)\nu(x) + \frac{\partial I(x)}{\partial t} \right)^2 + \lambda (\|\nabla(\nu_1(x))\|^2 + \|\nabla(\nu_2(x))\|^2) dx$$

MOTION ESTIMATION ALGORITHM

Three important elements

① Motion Models

- ① Spatial Motion Model
- ② Temporal Motion Model
- ③ Region of Support
- ④ Observation Model

② Estimation criteria

- ① Pixel-Domain Criteria
- ② Regularization

③ Search Strategies

- ① Matching
- ② Gradient-based techniques

SEARCH STRATEGIES

- Develop an efficient (complexity) and effective (solution quality) strategy for finding an estimate of motion parameters.

MATCHING

- For a small number of motion parameters
- A small state space for each of them
- Minimizing a prediction error
- Motion-compensated predictions $\tilde{I}_k[n]$ for various motion candidates \mathbf{d} are matched with $I_k[n]$ within the region of support of the motion model.
- The candidate yielding the best match for a given criterion becomes the optimal estimate.

GRADIENT-BASED TECHNIQUE

- Estimation criteria E is differentiable.
- To avoid non-linear optimization I is usually linearized using Taylor expansion with respect to $d[n]$.
- Gradient-based estimation yields accurate results only in regions of small motion.
- The approach fails if motion is large. This deficiency is usually compensated for by a hierarchical or multiresolution implementation.

GLOBAL MOTION ESTIMATION ALGORITHM



Motion Estimation

Optical Flow

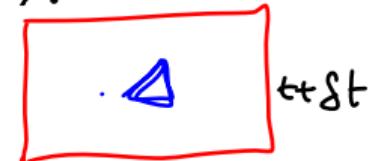
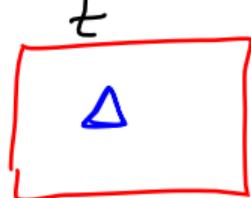
image sequence

• Motion field? image field

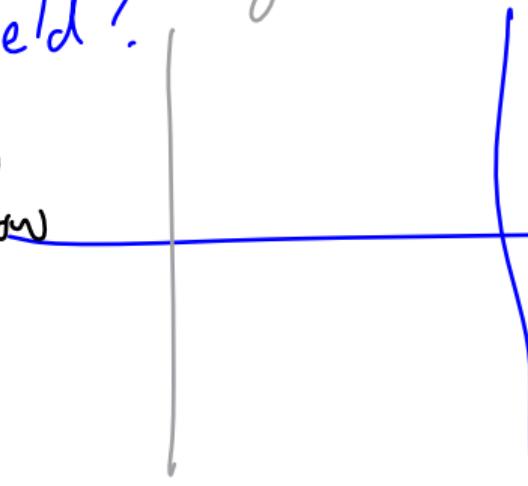
↓
optical flow

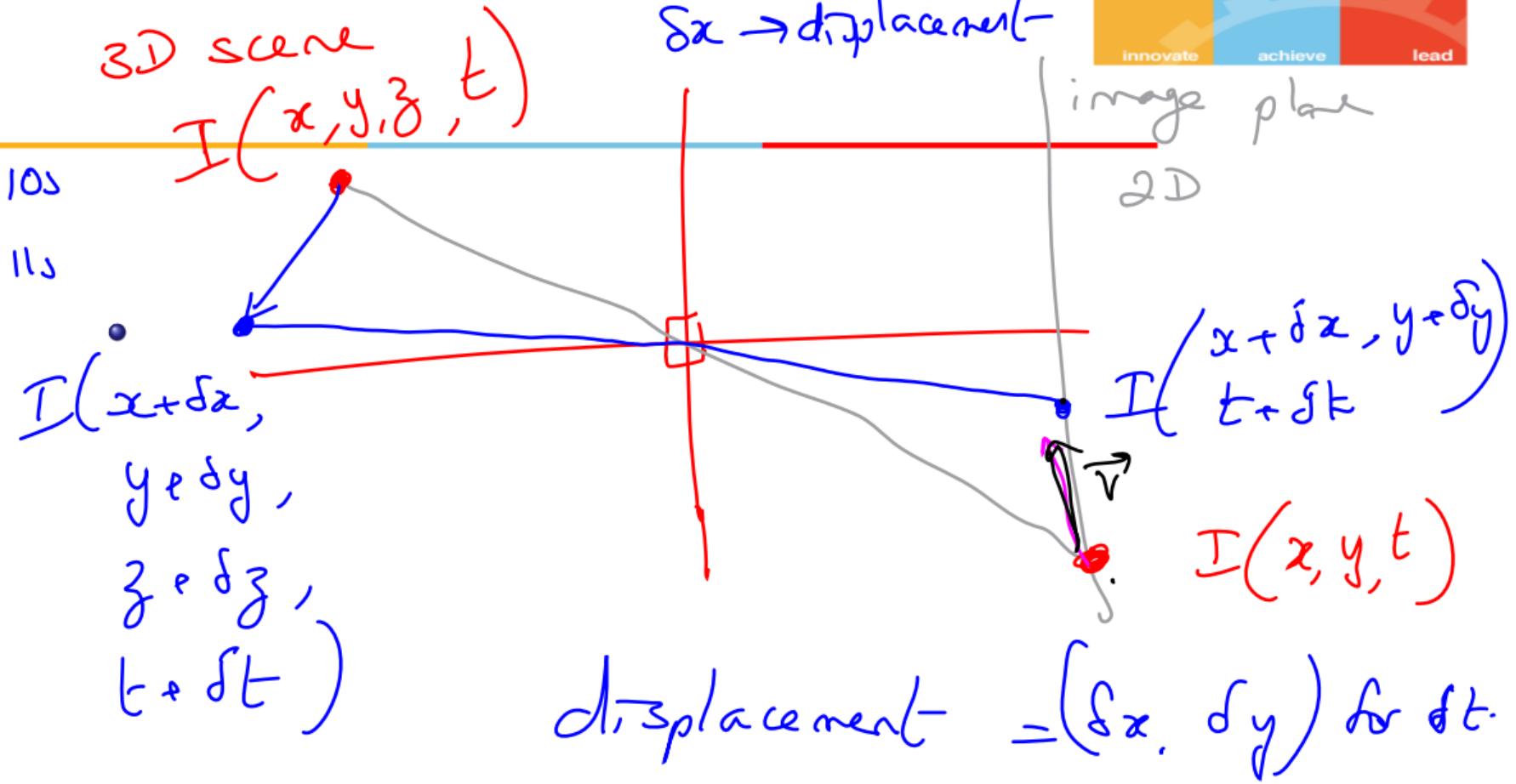


$$v = d/t$$
$$d = vt$$



motion field
in scene





obj is at a relxity

$$\textcircled{v} = \text{displacement / time}$$

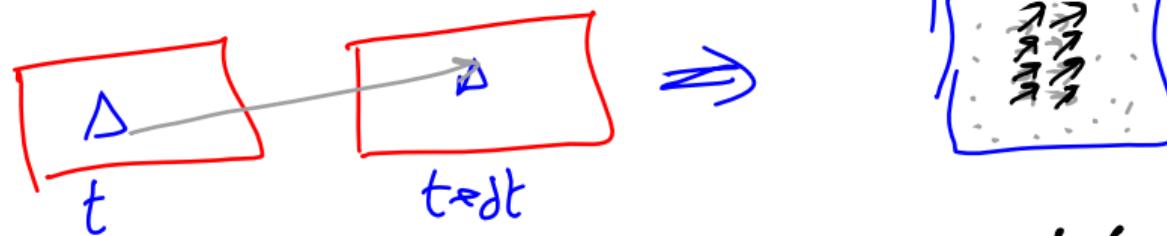
- distance in 2D x, y & direction of motion

$$\left\{ \begin{array}{l} \text{length of vector} = \text{distance pixel has} \\ \text{traversed} \\ \text{dischon of vector} = \text{velocity vector in } z \text{ direc} \\ u = \frac{s_x}{s.t} \end{array} \right.$$

velocity vector in y direction $v = \frac{\delta y}{dt}$

$\vec{v} = (u, v)$ = Motion vector of point P.

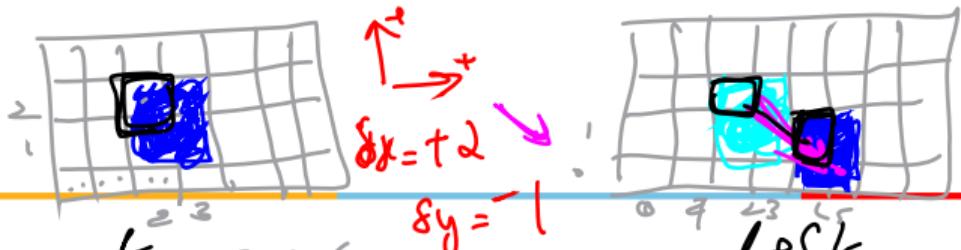
- All pixels together, motion field.



Measure the brightness patterns \approx intensity patterns.

(0,0,255)

(0,0,255)



t $(2, 1)$ $(2, 2)$
 $(3, 1)$ $(3, 2)$

$t + \delta t$

displacement $(\delta x, \delta y)$

video
108
30fps

$$\frac{f_1}{t} \quad \frac{f_2}{t + \delta t}$$

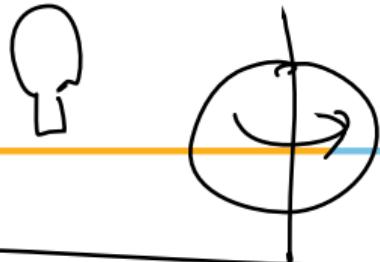
- $I(x, y, t)$ $I(x + \delta x, y + \delta y, t + \delta t)$

Assumption 1. \Rightarrow brightness/intensity remain constant

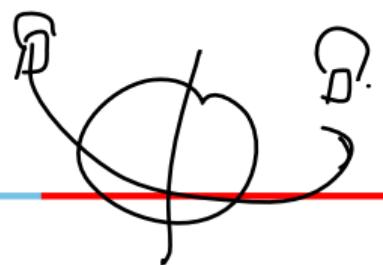
$$I(x + \delta x, y + \delta y + t + \delta t) = I(x, y, t) \rightarrow (1).$$

Assumption 2 $\delta t \rightarrow$ really small
 $\delta x, \delta y \rightarrow$ small ✓

33 ms/sec



M.F ✓
OF ✗



innovate achieve lead
 $M_F = 0$
OF ✓

- Taylor series expansion (first order)

$$f(x + \delta x) = f(x) + \frac{\partial f}{\partial x} \delta x \rightarrow 0 .$$

$$I(x + \delta x, y + \delta y, t + \delta t) = I(x, y, t) + \frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial t} \delta t$$

$$I(x + \delta x, y + \delta y + t + \delta t) = I(x, y, t) + I_x \delta x + I_y \delta y + I_t \delta t$$

$$I(x + \delta x, y + \delta y + t + \delta t) = I(x, y, t) \rightarrow (1)$$

(2) - (1)

$$I_x \delta x + I_y \delta y + I_t \delta t = 0$$

divide δt

$$\frac{I_x \delta x}{\delta t} + \frac{I_y \delta y}{\delta t} + \frac{I_t}{\delta t} = 0$$

limit $\delta t \rightarrow 0$

$$I_x u + I_y v + I_t = 0$$

$$(I_x, I_y, I_t)$$

2 frames finite differences.

(u, v) = velocity vector
optical flow



$$I_x u + I_y v + \frac{I}{t} = 0$$

optical
constraint
equation.
(u,v)

$$I_x = \frac{\partial I}{\partial x}$$

} spatial derivatives.

$$I_y = \frac{\partial I}{\partial y}$$

knows

$$I_t = \frac{\partial I}{\partial t}$$

temporal derivative.

$$u = \frac{\delta x}{\delta t}$$

} optical flow

$$v = \frac{\delta y}{\delta t}$$

} unknown









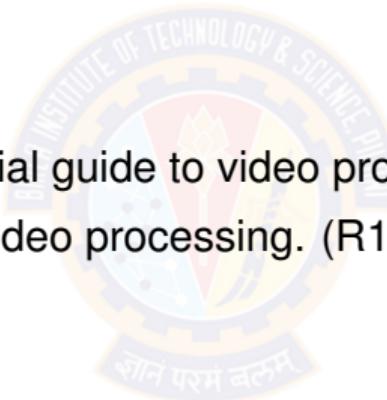






REFERENCES

- ① Bovik, Alan C. The essential guide to video processing. (T1) Ch 3
- ② Tekalp, A. Murat. Digital video processing. (R1)



Thank You!



VIDEO ANALYTICS MODULE # 2 : MOTION DETECTION AND ESTIMATION

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

DL Team, BITS Pilani

The instructor is gratefully acknowledging
the authors who made their course
materials freely available online.

This deck is prepared by Seetha Parameswaran.

TABLE OF CONTENTS

1 MODULE 2

2 INTRODUCTION TO MOTION

3 MATH PRELIMINARIES

4 MOTION DETECTION

MODULE TOPICS....

- MRF and MAP
- Motion detection
- Motion estimation
- Optical Flow Motion estimation
- MAP estimation for Dense motion
- Application

TABLE OF CONTENTS

1 MODULE 2

2 INTRODUCTION TO MOTION

3 MATH PRELIMINARIES

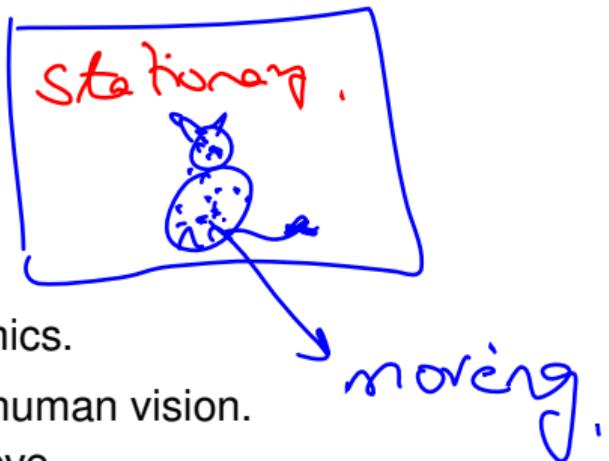
4 MOTION DETECTION

VIDEO AND MOTION

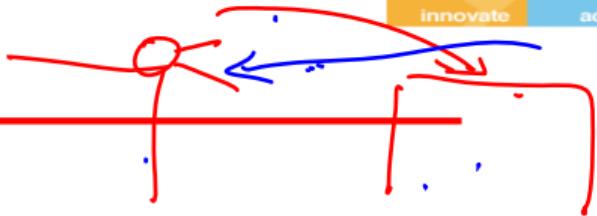
PIXELS. INTENSITY.

$I(x, y)$

- Video captures motion.
A single image provides snapshot of a scene.
A sequence of images records scene's dynamics.
- The recorded motion is a very strong cue for human vision.
Easy to recognize objects as soon as they move.

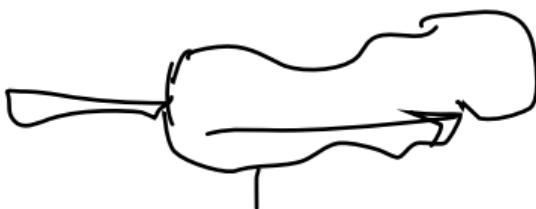


MOTION



- Motion is important for video processing and compression for two reasons.
 - ➊ Motion carries information about spatio-temporal relationships between objects in the field of view of a camera.
 - ★ used in applications such as traffic monitoring or security surveillance.
 - ➋ Image properties, such as intensity or color, have a very high correlation in the direction of motion. They do not change significantly when tracked over time
 - ★ used for the removal of temporal redundancy in video coding.
- Two-dimensional (2D) motion of intensity patterns in the image plane is referred to as **apparent motion**.

MOTION RELATED TASKS



MOTION DETECTION identify image points as moving or stationary.

MOTION ESTIMATION measure how image points move.

MOTION SEGMENTATION identify groups of image points moving similarly.

salt & pepper noise $\leftarrow g$

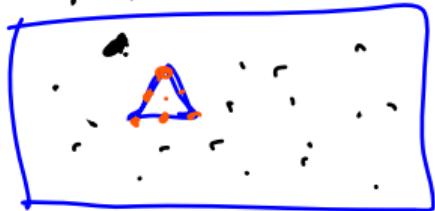


TABLE OF CONTENTS

~~analog
continuous~~

1 MODULE 2

30FPS



video .
~~10 sec~~
~~30x10 = 300~~ ~~1024 x 780~~ grid

300 frames. ↑

Sandu
my mobile .

2 INTRODUCTION TO MOTION

i-phone .

3 MATH PRELIMINARIES

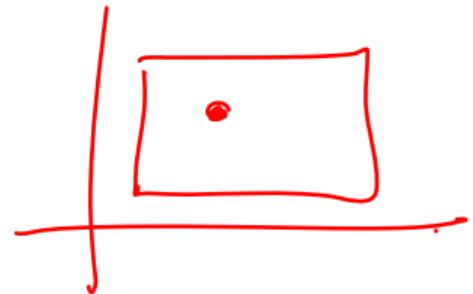
300 frames → 1600 x 1200
grid

4 MOTION DETECTION

IMAGE SEQUENCE

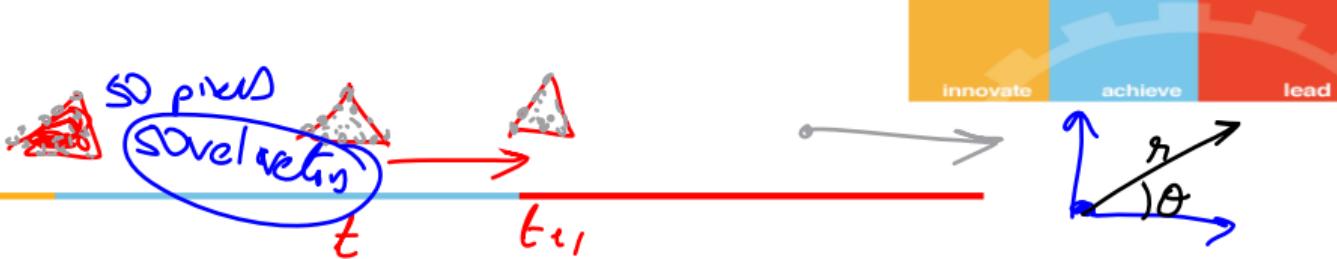
- Intensity of image sequence

$$\mathcal{I}: \Omega \times \mathcal{T} \rightarrow R^+$$



- Ω - spatial domain
- \mathcal{T} - temporal domain
- $\mathbf{x} = (x_1, x_2)^T \in \Omega$ - Spatial position of a point in the image sequence
- $t \in \mathcal{T}$ - Temporal position of a point in the image sequence

MOTION



- $\nu = (\nu_1, \nu_2)^T$ - Velocity vector to represent motion in continuous images.
- ν_t - Dense velocity field or motion field, that is, the set of all velocity vectors within the image, at time t .
- \mathbf{b}_t - small number of motion parameters. Reduce computational complexity.

$$\nu_t \xrightarrow{\text{transformation}} \mathbf{b}_t$$

- \mathbf{d} - Displacement / velocity vector to represent motion in discrete images

Set of pixels moving

CONTINUOUS VS DISCRETE REPRESENTATION

$$\left(\begin{pmatrix} x_1 & x_2 \end{pmatrix}^T, t \right) \quad \left(\begin{pmatrix} n_1 & n_2 \end{pmatrix}^T, t_k \right)$$

Representation	Continuous	Discrete
Time	t	t_k
Position	(\mathbf{x}, t)	$((n), t_k) = ((n_1, n_2)^T, t_k)$
Image	$I(\mathbf{x}, t)$	$I[n, t]$
Motion	$\mathbf{v} = (\nu_1, \nu_2)^T$	

BINARY HYPOTHESIS TESTING



- Let y be an observation.
- Let Y be the associated random variable.
- Two hypotheses *assumption*
 - H_0 - probability distributions $P(Y = y|H_0)$
 - H_1 - probability distributions $P(Y = y|H_1)$
- Goal: Decide from which of the two distributions a given y is selected.

BINARY HYPOTHESIS TESTING

- 4 possibilities for true hypothesis /decision
 - ▶ $H_0 \mid H_0$ - correct choice
 - ▶ $H_1 \mid H_1$ - correct choice
 - ▶ $H_0 \mid H_1$ - error
 - ▶ $H_1 \mid H_0$ - error
- To make a decision, a decision criterion is needed that attaches some relative importance to the four possibilities. i.e assign cost to each decision.

event = Hypothesis.

BAYES CRITERION

- Two a priori probabilities

A = H_0

► π_0 for H_0

► $\pi_1 = 1 - \pi_0$ for H_1

B = H_1

- Design a decision rule so that on average the cost associated with making a decision based on y is minimal.
- Optimal decision can be made according to the following rule

$$P(A|B)$$

$$= P(B|A) \frac{P(A)}{P(B)}$$

$$\frac{P_1}{P_0} = \frac{P(y = y | H_1)}{P(Y = y | H_0)}$$

Likelihood ratio

ϑ
constant

$\frac{\pi_0}{\pi_1}$
prior probability

posterior = likelihood * prior

$$P(H_0 | H_1)$$

= constant

$$\frac{P(H_0)}{P(H_1)}$$

$$P(H_1 | H_0)$$

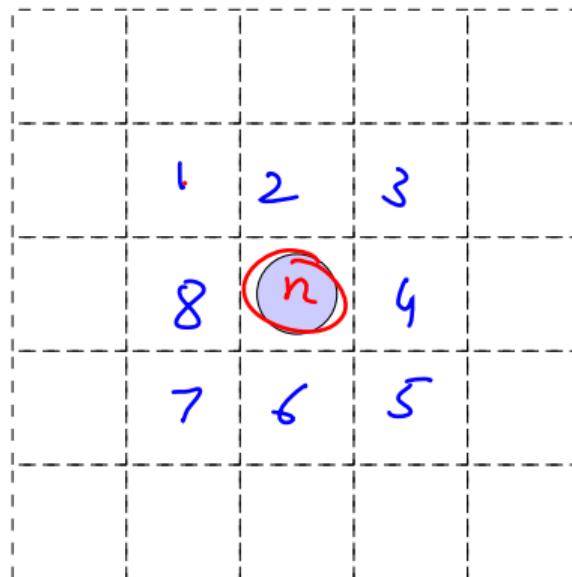
$$P(H_0 | H_1)$$

$$= P(H_1 | H_0) \cdot \frac{P(H_0)}{P(H_1)}$$

SAMPLING GRID

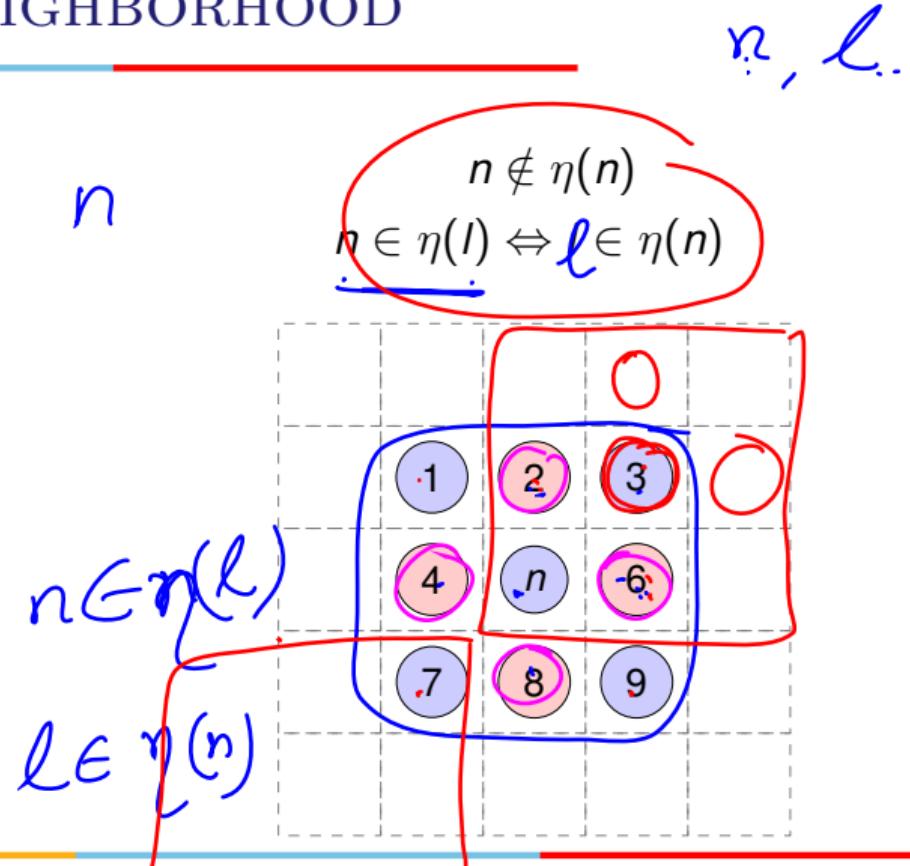
- Let Λ be a sampling grid in R^N .
- Let $n \in \Lambda$ be any point the grid.

date / point / pixel / image
point



SAMPLING GRID AND NEIGHBORHOOD

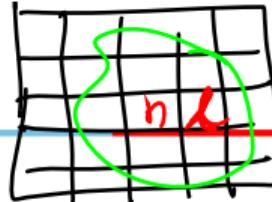
- Let $n \in \Lambda$ be any point the grid.
- Let $\eta(n)$ be a neighborhood of $n \in \Lambda$.
- The first-order neighborhood consists of immediate top, bottom, left, and right neighbors of n .
- Let \mathcal{N} be a neighborhood system, a collection of neighborhoods of all $n \in \Lambda$.



102wr740

RANDOM FIELD

0 - 1



RP $\rightarrow \Lambda$ (grid)

RV $\rightarrow n$ (pixel)

- A random field \mathcal{Y} over Λ is a multidimensional random process where each site $n \in \Lambda$ is assigned a random variable.
- A random field \mathcal{Y} with the following properties:

state $\rightarrow \Gamma \{s, m\}$

$$P(\mathcal{Y} = \nu) > 0 \quad \forall \nu \in \Gamma$$

$$P(\mathcal{Y}_n = \nu_n \mid \mathcal{Y}_l = \nu_l, \forall l \neq n) = P(\mathcal{Y}_n = \nu_n \mid \mathcal{Y}_l = \nu_l, \forall l \in \eta(n))$$

*n in grid
v in set
 $\forall n \in \Lambda, \forall \nu \in \Gamma$*

P is a probability measure, and is called a Markov random field with state space Γ .

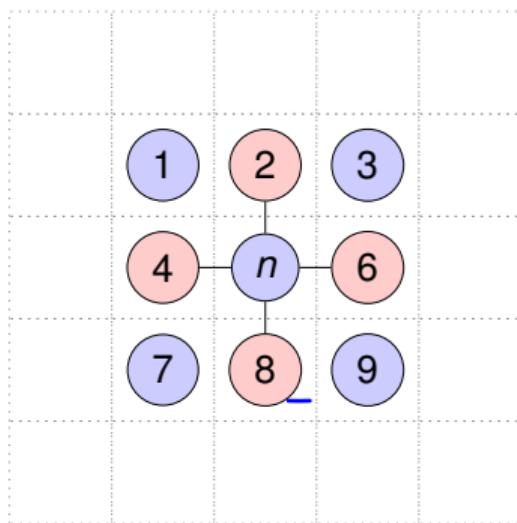
$$P(\mathcal{Y} \in s) > 0$$

$$P(l \in m) > 0$$

$s \in \eta(n)$
 $l \neq n$

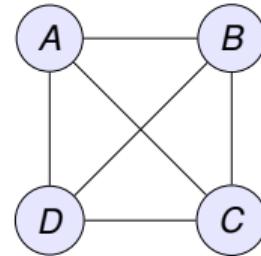
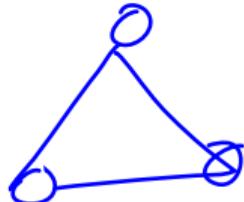
CLIQUE

- A clique c defined over Λ with respect to \mathcal{N} is a subset of Λ such that either c consists of a single site or every pair of sites in c are neighbors, that is, belong to η .
- The set of all cliques is denoted by C .
- A two-element spatial clique $\{n, l\}$ are two immediate horizontal, vertical or diagonal neighbors.

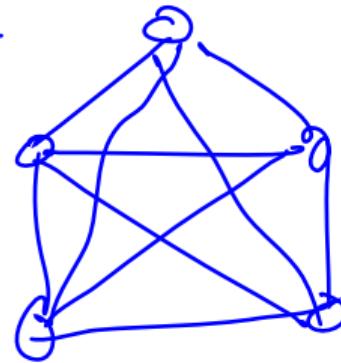


CLIQUE IN GENERAL

- Clique is a subset of vertices of an undirected graph such that every two distinct vertices in the clique are adjacent.
- A **clique** is a subset of nodes in which every node is connected to every other node.
- A **maximal clique** is a clique which cannot be extended by the addition of another node.



dense



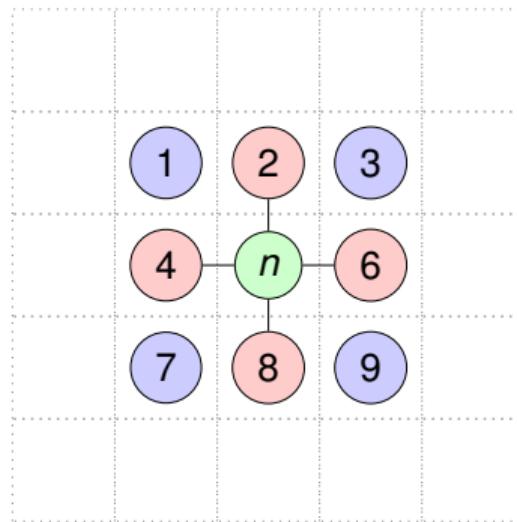
MARKOV RANDOM FIELD

- A Markov Random Field (MRF) is a graphical model of a joint probability distribution.
- Edges encode conditional independence.
- Let X_S be the set of random variables associated with the set of nodes S .
- Rule: Given disjoint subsets of nodes A , B and C , X_A is conditionally independent of X_B given X_C if there is no path from any node in A to any node in B that doesn't pass through a node of C . relationships.
- Markov property tells us that the joint distribution of X is determined entirely by the **local conditional distributions** $P(X_n \mid X_{\eta(n)})$

MARKOV RANDOM FIELD

Given the pink nodes, the green node is conditionally independent of all other nodes.

$$n \perp \{1, 3, 7, 9\} \mid \{2, 4, 6, 8\}$$



GIBBS DISTRIBUTION

- A Gibbs distribution on grid Λ and neighborhood \mathcal{N} takes the form:

$$P(\mathcal{Y} = \nu) = \frac{1}{Z} \exp\left(\frac{-1}{T} U(\nu)\right)$$

Z - Partition Function

T - temperature and is often taken to be 1.

U - energy function or potential function

$$U(\nu) = \sum_{c \in C} V(\nu, c)$$

For two element clique $\{n, l\}$ then $U(n, l) = V(\nu[n], \nu[l])$

MRF EQUIVALENCE TO GIBBS DISTRIBUTION

- The equivalence between Markov random fields and Gibbs distributions is provided through the important **Hammersley-Clifford theorem**.
- Theorem states that \mathcal{Y} is a MRF on Λ with respect to \mathcal{N} if and only if its probability distribution is a Gibbs distribution with respect to Λ and \mathcal{N} .

MAP ESTIMATION

- Let Y be a random field of observations
- Let \mathcal{Y} be a random field modeling the quantity we want to estimate based on Y
- .
- Let y, ν be their respective realizations.
- y could be a difference between two images.
- ν could be a field of motion detection labels.
- To compute ν based on y (MAP) estimation is

$$\hat{\nu} = \arg \max_{\nu} \underbrace{P(\mathcal{Y} = \nu \mid y)}_{\text{maximum of the posterior probability}} = \arg \max_{\nu} \underbrace{P(\mathcal{Y} = y \mid \nu) \cdot P(\mathcal{Y} = \nu)}_{\text{maximum likelihood}}$$

ML ESTIMATION

- To compute ν based on y (MAP) estimation is

$$\hat{\nu} = \arg \max_{\nu} \underbrace{P(\mathcal{Y} = \nu \mid y)}_{\text{maximum of the posterior probability}} = \arg \max_{\nu} \underbrace{P(\mathcal{Y} = y \mid \nu) \cdot P(\mathcal{Y} = \nu)}_{\text{maximum likelihood}}$$

- If $P(\mathcal{Y} = \nu)$ is the same for all realizations ν , then only the likelihood $P(\mathcal{Y} = y \mid \nu)$ is maximized, resulting in the maximum likelihood (ML) estimation.

TABLE OF CONTENTS

1 MODULE 2

2 INTRODUCTION TO MOTION

3 MATH PRELIMINARIES

4 MOTION DETECTION

GOAL OF MOTION DETECTION

- Identify which image points, or, regions of the image, have moved.
- Motion detection applies to images acquired with a **static camera**.
- Motion of image points is not perceived directly but rather through intensity changes.
- Such intensity changes over time may be also induced by camera noise or illumination variations.

HYPOTHESIS TESTING

- Simplest motion detection algorithms
- Let H_S and H_M be two hypotheses.
- H_S declaring an image point at n as stationary (S). State S means 0.
- H_M says an image point at n is moving (M). State M means 1.
- Let q be noise. Noise is assumed as zero-mean Gaussian with variance σ^2 in stationary areas and uniformly distributed in range $[-L, L]$ in moving areas.
- Assume $I_k[n] = I_{k-1}[n] + q$.
- P_S is assumed Gaussian, while P_M is assumed uniform.
- The motivation is that in stationary areas only camera noise will distinguish same-position pixels, whereas in moving areas this difference is attributed to motion and therefore unpredictable.

HYPOTHESIS TESTING WITH FIXED THRESHOLD

- Let an observation, upon which we intend to select one of the two hypotheses be

$$\rho_k[n] = I_k[n] - I_{k-1}[n]$$

- Hypothesis test is:

$$\rho_k^2[n] > \theta \quad \forall n \in M$$

$$\rho_k^2[n] < \theta \quad \forall n \in S$$

$$\theta = 2\sigma^2 \ln \left(\vartheta \cdot 2L \cdot P_S / (\sqrt{2\pi\sigma^2} \cdot P_M) \right)$$

- Not robust to noise in the image
- For small θ "noisy" detection masks result
- For large θ only object boundaries and its most textured parts are detected.

HYPOTHESIS TESTING, FIXED THRESHOLD, AVERAGING

- Averaging the observations over an N -point spatial window W_n centered at n

$$\frac{1}{N} \sum_{m \in W_n} \rho_k^2[n] > \theta \quad \forall n \in M$$

- Used to attenuate the impact of noise

MOTION DETECTION BASED ON FRAME DIFFERENCES

- Motion detection based on frame differences (last 3 slides) does not perform well for large, untextured objects (e.g., a large, uniformly colored truck).
- Only pixels n where $|I_k[n] - I_{k-1}[n]|$ is sufficiently large can be reliably detected.
- Such pixels concentrate in narrow areas close to moving boundaries where object intensity is distinct from the background in the previous frame.

MOTION DETECTION BY COMPARING BACKGROUND INTENSITY

- Comparing the current intensity $I_k[n]$ to background intensity $B_k[n]$ instead of the previous frame $I_{k-1}[n]$.

$$\rho_k[n] = I_k[n] - B_k[n]$$

- Estimate $B_k[n]$ by means of temporal averaging or median filtering the intensity at each n .
- Median can suppress intensities associated with moving objects (large window) and is fast. It fails in the presence of parasitic motion, such as fluttering leaves or waves on water surface.

HYPOTHESIS TESTING WITH FIXED THRESHOLD

Non-parametric distributions

- At each location n of frame k , an estimate of the stationary (background) probability distribution is computed from K recent frames as

$$P_S(I_k[n]) = \frac{1}{K} \sum_{i=1}^K \kappa(I_k[n] - I_{k-i}[n])$$

κ is a zero-mean Gaussian with variance σ^2 considered as constant

- Hypothesis test:

$$P_S(I_k[n]) > \theta \quad \forall n \in S$$

$$P_S(I_k[n]) < \theta \quad \forall n \in M$$

$$\theta = \frac{P_M}{2L\vartheta P_S}$$

MAP MRF FORMULATION

- To find a MAP estimate of the random field E_k , maximize the posterior probability

$$P(E_k = e_k \mid \rho_k)$$

- Let $|I_k[n] - I_{k-1}[n]|$ be an observation modelled as

$$\rho_k[n] = \xi(e_k[n]) + q[n]$$

where q is zero-mean uncorrelated Gaussian noise with variance σ^2

$$\xi(e_k[n]) = \begin{cases} 0 & \text{if } e_k[n] = \mathcal{S} \\ \alpha & \text{if } e_k[n] = \mathcal{M} \end{cases}$$

α is average temporal intensity difference based on previous-time moving labels e_{k-1}

MAP MRF FORMULATION

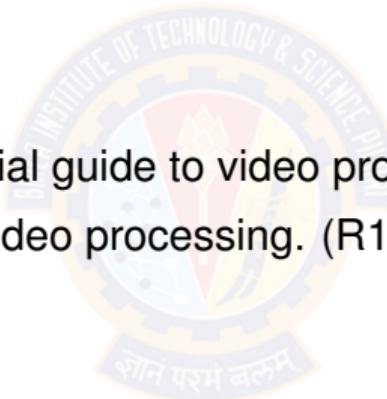
- ξ attempts to closely model the observations since for a static image point it is zero, whereas for a moving point it tracks average temporal intensity mismatch.
- Likelihood $P(R_k = \rho_k | e_k)$
- Gibbs distribution for the a priori probability $\pi(E_k = e_k)$
- The overall energy function





REFERENCES

- ① Bovik, Alan C. The essential guide to video processing. (T1) Ch 3
- ② Tekalp, A. Murat. Digital video processing. (R1) Ch 1.2



Thank You!



VIDEO ANALYTICS MODULE # 2 : MOTION DETECTION AND ESTIMATION

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

DL Team, BITS Pilani

The instructor is gratefully acknowledging
the authors who made their course
materials freely available online.

This deck is prepared by Seetha Parameswaran.

TABLE OF CONTENTS

1 MODULE 2

2 INTRODUCTION TO MOTION

3 MATH PRELIMINARIES

4 MOTION DETECTION

MODULE TOPICS....

- MRF and MAP
- Motion detection
- Motion estimation
- Optical Flow Motion estimation
- MAP estimation for Dense motion
- Application

TABLE OF CONTENTS

1 MODULE 2

2 INTRODUCTION TO MOTION

3 MATH PRELIMINARIES

4 MOTION DETECTION

VIDEO AND MOTION

- Video captures motion.
 - A single image provides snapshot of a scene.
 - A sequence of images records scene's dynamics.
- The recorded motion is a very strong cue for human vision.
 - Easy to recognize objects as soon as they move.

MOTION

- Motion is important for video processing and compression for two reasons.
 - ① Motion carries information about spatio-temporal relationships between objects in the field of view of a camera.
 - ★ used in applications such as traffic monitoring or security surveillance.
 - ② Image properties, such as intensity or color, have a very high correlation in the direction of motion. They do not change significantly when tracked over time
 - ★ used for the removal of temporal redundancy in video coding.
- Two-dimensional (2D) motion of intensity patterns in the image plane is referred to as **apparent motion**.

MOTION RELATED TASKS

MOTION DETECTION identify image points as moving or stationary.

MOTION ESTIMATION measure how image points move.

MOTION SEGMENTATION identify groups of image points moving similarly.

TABLE OF CONTENTS

1 MODULE 2

2 INTRODUCTION TO MOTION

3 MATH PRELIMINARIES

4 MOTION DETECTION

IMAGE SEQUENCE

- Intensity of image sequence

$$\mathcal{I} : \Omega \times \mathcal{T} \rightarrow R^+$$

- Ω - spatial domain
- \mathcal{T} - temporal domain
- $\mathbf{x} = (x_1, x_2)^T \in \Omega$ - Spatial position of a point in the image sequence
- $t \in \mathcal{T}$ - Temporal position of a point in the image sequence

MOTION

- $\nu = (\nu_1, \nu_2)^T$ - Velocity vector to represent motion in continuous images.
- ν_t - Dense velocity field or motion field, that is, the set of all velocity vectors within the image, at time t .
- \mathbf{b}_t - small number of motion parameters. Reduce computational complexity.

$$\nu_t \xrightarrow{\text{transformation}} \mathbf{b}_t$$

- \mathbf{d} - Displacement / velocity vector to represent motion in discrete images

CONTINUOUS VS DISCRETE REPRESENTATION

Representation	Continuous	Discrete
Time	t	t_k
Position	(\mathbf{x}, t)	$((n), t_k) = ((n_1, n_2)^T, t_k)$
Image	$I(\mathbf{x}, t)$	$I[\mathbf{n}, t]$
Motion	$\mathbf{v} = (\nu_1, \nu_2)^T$	

BINARY HYPOTHESIS TESTING

- Let y be an observation.
- Let Y be the associated random variable.
- Two hypotheses
 - ▶ H_0 - probability distributions $P(Y = y|H_0)$
 - ▶ H_1 - probability distributions $P(Y = y|H_1)$
- Goal: **Decide from which of the two distributions a given y is selected.**

BINARY HYPOTHESIS TESTING

- 4 possibilities for true hypothesis /decision
 - ▶ $H_0 \mid H_0$ - correct choice
 - ▶ $H_1 \mid H_1$ - correct choice
 - ▶ $H_0 \mid H_1$ - error
 - ▶ $H_1 \mid H_0$ - error
- To make a decision, a decision criterion is needed that attaches some relative importance to the four possibilities. i.e assign cost to each decision.

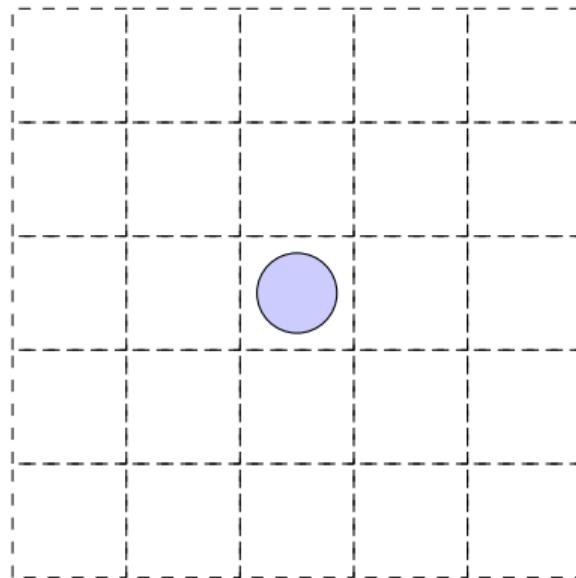
BAYES CRITERION

- Two a priori probabilities
 - ▶ π_0 for H_0
 - ▶ $\pi_1 = 1 - \pi_0$ for H_1
- Design a decision rule so that on average the cost associated with making a decision based on y is minimal.
- Optimal decision can be made according to the following rule

$$\frac{P_1}{P_0} = \underbrace{\frac{P(y = y | H_1)}{P(Y = y | H_0)}}_{\text{Likelihood ratio}} \gtrless \underbrace{\vartheta}_{\text{constant}} \quad \underbrace{\frac{\pi_0}{\pi_1}}_{\text{prior probability}}$$

SAMPLING GRID

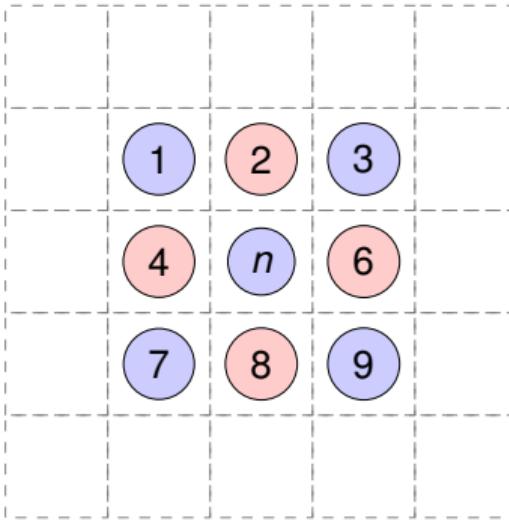
- Let Λ be a sampling grid in R^N .
- Let $n \in \Lambda$ be any point the grid.



SAMPLING GRID AND NEIGHBORHOOD

- Let $n \in \Lambda$ be any point the grid.
- Let $\eta(n)$ be a neighborhood of $n \in \Lambda$.
- The first-order neighborhood consists of immediate top, bottom, left, and right neighbors of n .
- Let \mathcal{N} be a neighborhood system, a collection of neighborhoods of all $n \in \Lambda$.

$$\begin{aligned} n &\notin \eta(n) \\ n \in \eta(l) &\Leftrightarrow l \in \eta(n) \end{aligned}$$



RANDOM FIELD

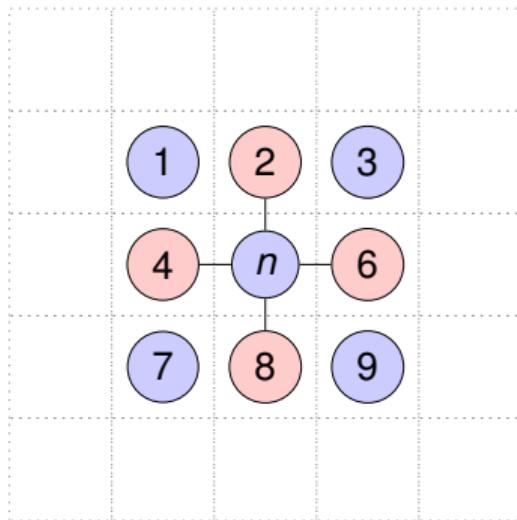
- A random field \mathcal{Y} over Λ is a multidimensional random process where each site $n \in \Lambda$ is assigned a random variable.
- A random field \mathcal{Y} with the following properties:

$$\left\{ \begin{array}{l} P(\mathcal{Y} = \nu) > 0 \quad \forall \nu \in \Gamma \\ P(\mathcal{Y}_n = \nu_n \mid \mathcal{Y}_I = \nu_I, \forall I \neq n) = P(\mathcal{Y}_n = \nu_n \mid \mathcal{Y}_I = \nu_I, \forall I \in \eta(n)) \quad \forall n \in \Lambda, \forall \nu \in \Gamma \end{array} \right.$$

P is a probability measure, and is called a Markov random field with state space Γ .

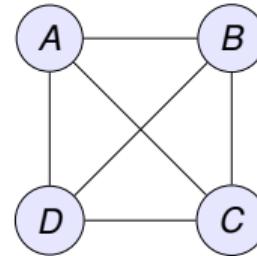
CLIQUE

- A clique c defined over Λ with respect to \mathcal{N} is a subset of Λ such that either c consists of a single site or every pair of sites in c are neighbors, that is, belong to η .
- The set of all cliques is denoted by C .
- A two-element spatial clique $\{n, l\}$ are two immediate horizontal, vertical or diagonal neighbors.



CLIQUE IN GENERAL

- Clique is a subset of vertices of an undirected graph such that every two distinct vertices in the clique are adjacent.
- A **clique** is a subset of nodes in which every node is connected to every other node.
- A **maximal clique** is a clique which cannot be extended by the addition of another node.



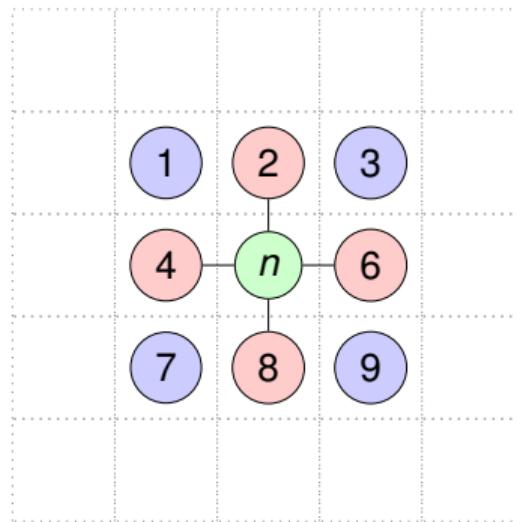
MARKOV RANDOM FIELD

- A Markov Random Field (MRF) is a graphical model of a joint probability distribution.
- Edges encode conditional independence.
- Let X_S be the set of random variables associated with the set of nodes S .
- Rule: Given disjoint subsets of nodes A , B and C , X_A is conditionally independent of X_B given X_C if there is no path from any node in A to any node in B that doesn't pass through a node of C . relationships.
- Markov property tells us that the joint distribution of X is determined entirely by the **local conditional distributions** $P(X_n \mid X_{\eta(n)})$

MARKOV RANDOM FIELD

Given the pink nodes, the green node is conditionally independent of all other nodes.

$$n \perp \{1, 3, 7, 9\} \mid \{2, 4, 6, 8\}$$



GIBBS DISTRIBUTION

- A Gibbs distribution on grid Λ and neighborhood \mathcal{N} takes the form:

$$P(\mathcal{Y} = \nu) = \frac{1}{Z} \exp\left(\frac{-1}{T} U(\nu)\right)$$

Z - Partition Function

T - temperature and is often taken to be 1.

U - energy function or potential function

$$U(\nu) = \sum_{c \in C} V(\nu, c)$$

For two element clique $\{n, l\}$ then $U(n, l) = V(\nu[n], \nu[l])$

MRF EQUIVALENCE TO GIBBS DISTRIBUTION

- The equivalence between Markov random fields and Gibbs distributions is provided through the important **Hammersley-Clifford theorem**.
- Theorem states that \mathcal{Y} is a MRF on Λ with respect to \mathcal{N} if and only if its probability distribution is a Gibbs distribution with respect to Λ and \mathcal{N} .

MAP ESTIMATION

- Let Y be a random field of observations
- Let \mathcal{Y} be a random field modeling the quantity we want to estimate based on Y
- .
- Let y, ν be their respective realizations.
- y could be a difference between two images.
- ν could be a field of motion detection labels.
- To compute ν based on y (MAP) estimation is

$$\hat{\nu} = \arg \max_{\nu} \underbrace{P(\mathcal{Y} = \nu \mid y)}_{\text{maximum of the posterior probability}} = \arg \max_{\nu} \underbrace{P(\mathcal{Y} = y \mid \nu) \cdot P(\mathcal{Y} = \nu)}_{\text{maximum likelihood}}$$

ML ESTIMATION

- To compute ν based on y (MAP) estimation is

$$\hat{\nu} = \arg \max_{\nu} \underbrace{P(\mathcal{Y} = \nu \mid y)}_{\text{maximum of the posterior probability}} = \arg \max_{\nu} \underbrace{P(\mathcal{Y} = y \mid \nu) \cdot P(\mathcal{Y} = \nu)}_{\text{maximum likelihood}}$$

- If $P(\mathcal{Y} = \nu)$ is the same for all realizations ν , then only the likelihood $P(\mathcal{Y} = y \mid \nu)$ is maximized, resulting in the maximum likelihood (ML) estimation.

TABLE OF CONTENTS

1 MODULE 2

2 INTRODUCTION TO MOTION

3 MATH PRELIMINARIES

4 MOTION DETECTION

ℓ, n - pixel .
 Λ - sampling grid
 $\eta(n)$ - neighbourhood
 N - set of η
 $n \in \eta(\ell) \Leftrightarrow \ell \in \eta(n)$.
 Γ - state space
 y - MRF-

GOAL OF MOTION DETECTION

$$g(x, y, t)$$

- Identify which image points, or, regions of the image, have moved.
- Motion detection applies to images acquired with a static camera.
- Motion of image points is not perceived directly but rather through intensity changes.
- Such intensity changes over time may be also induced by camera noise or illumination variations.

prev
 $t-1$

curr
 t

next
 $t+1$

frame

HYPOTHESIS TESTING

noise $q \rightarrow s$ Gaussian $g(0, \sigma^2)$

$\rightarrow m$ uniform ($= 1, \dots$)

- Simplest motion detection algorithms
 - Let H_S and H_M be two hypotheses.
 - H_S declaring an image point at n as stationary (S). State S means 0.
 - H_M says an image point at n is moving (M). State M means 1.
 - Let q be noise. Noise is assumed as zero-mean Gaussian with variance σ^2 in stationary areas and uniformly distributed in range $[-L, L]$ in moving areas.
 - Assume $I_k[n] = I_{k-1}[n] + q$.
 - P_S is assumed Gaussian, while P_M is assumed uniform.
 - The motivation is that in stationary areas only camera noise will distinguish same-position pixels, whereas in moving areas this difference is attributed to motion and therefore unpredictable.

$$\begin{array}{|c|c|c|} \hline 4 & 4 & 3 \\ \hline 5 & 5 & 2 \\ \hline 5 & 7 & 3 \\ \hline \end{array} + 2 = \begin{array}{|c|c|c|} \hline 6 & 4 & 4 \\ \hline 5 & 7 & 7 \\ \hline 5 & 7 & 117 \\ \hline \end{array}$$

$$I_k[n] = I[n] \text{ assumed uniform.}$$

HYPOTHESIS TESTING WITH FIXED THRESHOLD

- Let an observation, upon which we intend to select one of the two hypotheses be

MASK

- Hypothesis test is:

frame
difference

$$\begin{cases} \rho_k^2[n] > \theta & \forall n \in M \\ \rho_k^2[n] < \theta & \forall n \in S \end{cases}$$

$$\theta = 2\sigma^2 \ln \left(\vartheta \cdot 2L \cdot P_S / (\sqrt{2\pi}\sigma^2 \cdot P_M) \right)$$

$$\underline{\rho_k[n]} = I_k[n] - I_{k-1}[n]$$

n pixel

$$\theta = 2\sigma^2 \ln \left(\frac{\vartheta \cdot 2L \cdot P_S}{\sqrt{2\pi}\sigma^2 \cdot P_M} \right)$$

frame

- Not robust to noise in the image
- For small θ "noisy" detection masks result
- For large θ only object boundaries and its most textured parts are detected.

3	2	1
3	2	2
3	2	2

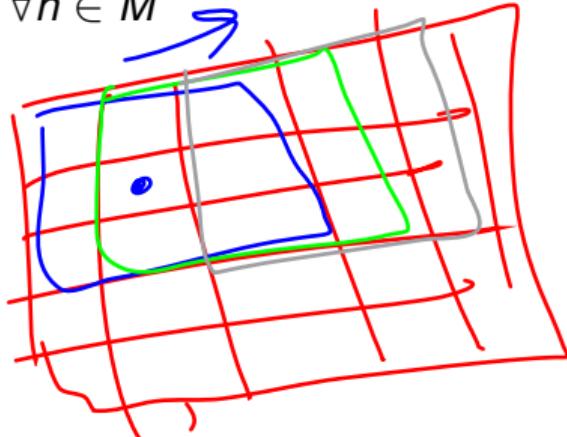
8	2	1
3	2	48
3	48	49

HYPOTHESIS TESTING, FIXED THRESHOLD, AVERAGING

- Averaging the observations over an N -point spatial window W_n centered at n

$$\frac{1}{N} \sum_{m \in W_n} \rho_k^2[m] > \theta \quad \forall n \in M$$

- Used to attenuate the impact of noise



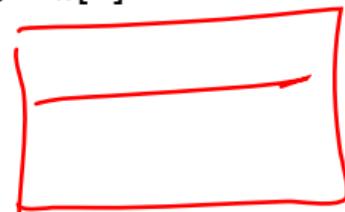
MOTION DETECTION BASED ON FRAME DIFFERENCES

- Motion detection based on frame differences (last 3 ~~slides~~) does not perform well for large, untextured objects (e.g., a large, uniformly colored truck).
- Only pixels n where $|I_k[n] - I_{k-1}[n]|$ is sufficiently large can be reliably detected.
- Such pixels concentrate in narrow areas close to moving boundaries where object intensity is distinct from the background in the previous frame.

MOTION DETECTION BY COMPARING BACKGROUND INTENSITY

- Comparing the current intensity $I_k[n]$ to background intensity $B_k[n]$ instead of the previous frame $I_{k-1}[n]$.

$$\rho_k[n] = I_k[n] - B_k[n]$$



- Estimate $B_k[n]$ by means of temporal averaging or median filtering the intensity at each n .
- Median can suppress intensities associated with moving objects (large window) and is fast. It fails in the presence of parasitic motion, such as fluttering leaves or waves on water surface.

HYPOTHESIS TESTING WITH FIXED THRESHOLD

Non-parametric distributions

- At each location n of frame k , an estimate of the stationary (background) probability distribution is computed from K recent frames as

$$P_S(I_k[n]) = \frac{1}{K} \sum_{i=1}^K \kappa(I_k[n] - I_{k-i}[n])$$

κ is a zero-mean Gaussian with variance σ^2 considered as constant

- Hypothesis test:

$$P_S(I_k[n]) > \theta \quad \forall n \in S$$

$$P_S(I_k[n]) < \theta \quad \forall n \in M$$

$$\theta = \frac{P_M}{2L\vartheta P_S}$$

MAP MRF FORMULATION

- To find a MAP estimate of the random field E_k , maximize the posterior probability

$$P(\underline{E}_k = \underline{e}_k | \rho_k)$$

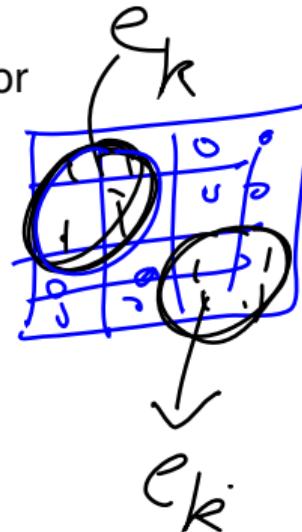
- Let $|I_k[n] - I_{k-1}[n]|$ be an observation modelled as

$$\rho_k[n] = \xi(e_k[n]) + q[n]$$

where q is zero-mean uncorrelated Gaussian noise with variance σ^2

$$\xi(e_k[n]) = \begin{cases} 0 & \text{if } e_k[n] = \mathcal{S} \\ \alpha & \text{if } e_k[n] = \mathcal{M} \end{cases}$$

α is average temporal intensity difference based on previous-time moving labels e_{k-1}



MAP MRF FORMULATION

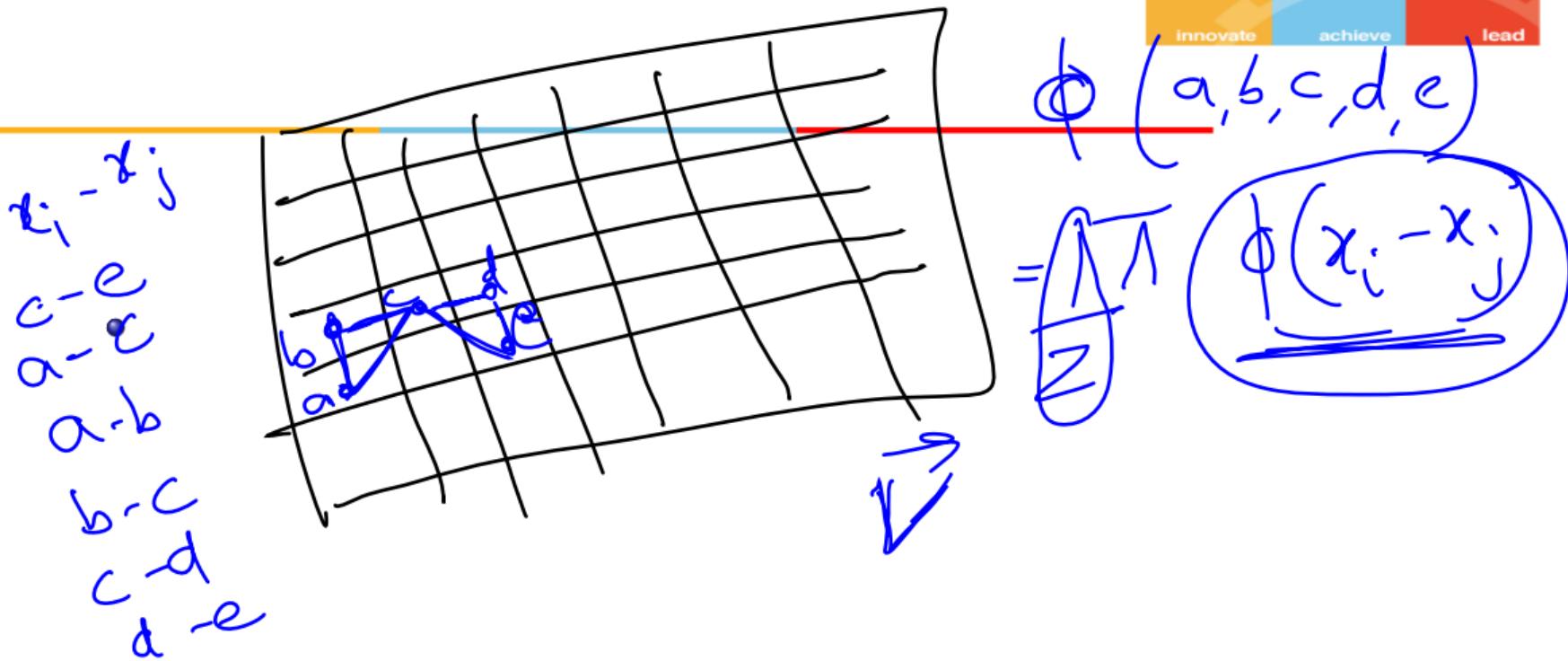
$$\text{energy} = \left(\rho_k - \xi(e_k) \right)^2 + \sum_{n, l \in \text{ec}} V(e_k^{[n]}, e_k^{[l]})$$

for

- ξ attempts to closely model the observations since for a static image point it is zero, whereas for a moving point it tracks average temporal intensity mismatch.
- Likelihood $P(R_k = \rho_k | e_k)$
- Gibbs distribution for the a priori probability $\pi(E_k = e_k)$
- The overall energy function

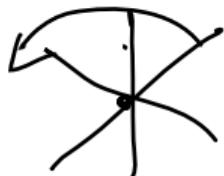
$$\sum_{k \in k-1} V(e_k^{[n]}, e_{k-1}^{[n]})$$

$$\text{Gibbs } \pi(E_k = e_k) =$$



Γ_{k-1}

3	3	3	3
3	3	3	255
3	3	255	3
3	255	3	3



Γ_k .

3	3	3	3
3	3	255	3
3	3	255	3
3	255	3	3

$$\left| \Gamma_R - \Gamma_{k-1} \right| =$$

$$g(e_k) =$$

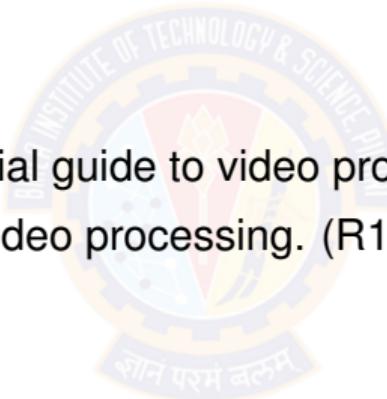
0	0	0	0
0	0	252	+252
0	0	0	0
0	252	+252	0

$$0 \quad n \in S$$

$$252 \quad n \in M$$

REFERENCES

- ① Bovik, Alan C. The essential guide to video processing. (T1) Ch 3
- ② Tekalp, A. Murat. Digital video processing. (R1) Ch 1.2



Thank You!



VIDEO ANALYTICS MODULE # 2 : MOTION DETECTION AND ESTIMATION

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

DL Team, BITS Pilani

The instructor is gratefully acknowledging
the authors who made their course
materials freely available online.

This deck is prepared by Seetha Parameswaran.

TABLE OF CONTENTS

- 1 MODULE 2
- 2 INTRODUCTION TO MOTION
- 3 MATH PRELIMINARIES
- 4 MOTION DETECTION
- 5 MOTION ESTIMATION

MODULE TOPICS....

- MRF and MAP
- Motion detection
- Motion estimation
- Optical Flow Motion estimation
- MAP estimation for Dense motion
- Application

TABLE OF CONTENTS

1 MODULE 2

2 INTRODUCTION TO MOTION

3 MATH PRELIMINARIES

4 MOTION DETECTION

5 MOTION ESTIMATION

VIDEO AND MOTION

- Video captures motion.
 - A single image provides snapshot of a scene.
 - A sequence of images records scene's dynamics.
- The recorded motion is a very strong cue for human vision.
 - Easy to recognize objects as soon as they move.

MOTION

- Motion is important for video processing and compression for two reasons.
 - ① Motion carries information about spatio-temporal relationships between objects in the field of view of a camera.
 - ★ used in applications such as traffic monitoring or security surveillance.
 - ② Image properties, such as intensity or color, have a very high correlation in the direction of motion. They do not change significantly when tracked over time
 - ★ used for the removal of temporal redundancy in video coding.
- Two-dimensional (2D) motion of intensity patterns in the image plane is referred to as **apparent motion**.

MOTION RELATED TASKS

MOTION DETECTION identify image points as moving or stationary.

MOTION ESTIMATION measure how image points move.

MOTION SEGMENTATION identify groups of image points moving similarly.

TABLE OF CONTENTS

1 MODULE 2

2 INTRODUCTION TO MOTION

3 MATH PRELIMINARIES

4 MOTION DETECTION

5 MOTION ESTIMATION

IMAGE SEQUENCE

- Intensity of image sequence

$$\mathcal{I} : \Omega \times \mathcal{T} \rightarrow R^+$$

- Ω - spatial domain
- \mathcal{T} - temporal domain
- $\mathbf{x} = (x_1, x_2)^T \in \Omega$ - Spatial position of a point in the image sequence
- $t \in \mathcal{T}$ - Temporal position of a point in the image sequence

MOTION

- $\nu = (\nu_1, \nu_2)^T$ - Velocity vector to represent motion in continuous images.
- ν_t - Dense velocity field or motion field, that is, the set of all velocity vectors within the image, at time t .
- \mathbf{b}_t - small number of motion parameters. Reduce computational complexity.

$$\nu_t \xrightarrow{\text{transformation}} \mathbf{b}_t$$

- \mathbf{d} - Displacement / velocity vector to represent motion in discrete images

CONTINUOUS VS DISCRETE REPRESENTATION

Representation	Continuous	Discrete
Time	t	t_k
Position	(\mathbf{x}, t)	$((n), t_k) = ((n_1, n_2)^T, t_k)$
Image	$I(\mathbf{x}, t)$	$I[\mathbf{n}, t]$
Motion	$\mathbf{v} = (\nu_1, \nu_2)^T$	

BINARY HYPOTHESIS TESTING

- Let y be an observation.
- Let Y be the associated random variable.
- Two hypotheses
 - ▶ H_0 - probability distributions $P(Y = y|H_0)$
 - ▶ H_1 - probability distributions $P(Y = y|H_1)$
- Goal: **Decide from which of the two distributions a given y is selected.**

BINARY HYPOTHESIS TESTING

- 4 possibilities for true hypothesis /decision
 - ▶ $H_0 \mid H_0$ - correct choice
 - ▶ $H_1 \mid H_1$ - correct choice
 - ▶ $H_0 \mid H_1$ - error
 - ▶ $H_1 \mid H_0$ - error
- To make a decision, a decision criterion is needed that attaches some relative importance to the four possibilities. i.e assign cost to each decision.

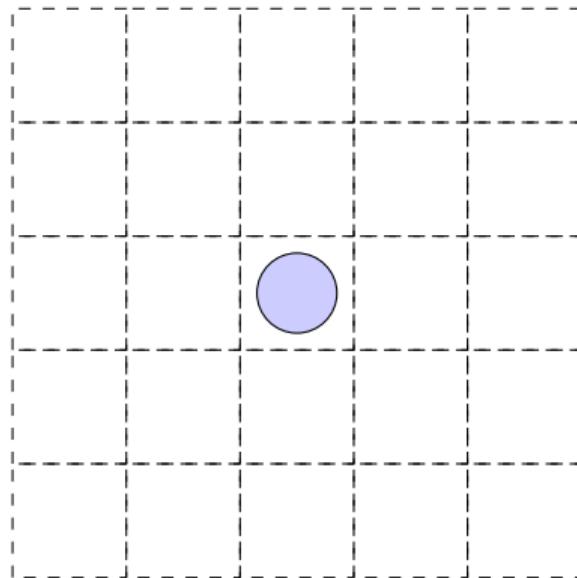
BAYES CRITERION

- Two a priori probabilities
 - ▶ π_0 for H_0
 - ▶ $\pi_1 = 1 - \pi_0$ for H_1
- Design a decision rule so that on average the cost associated with making a decision based on y is minimal.
- Optimal decision can be made according to the following rule

$$\frac{P_1}{P_0} = \underbrace{\frac{P(y = y | H_1)}{P(Y = y | H_0)}}_{\text{Likelihood ratio}} \gtrless \underbrace{\vartheta}_{\text{constant}} \quad \underbrace{\frac{\pi_0}{\pi_1}}_{\text{prior probability}}$$

SAMPLING GRID

- Let Λ be a sampling grid in R^N .
- Let $n \in \Lambda$ be any point the grid.

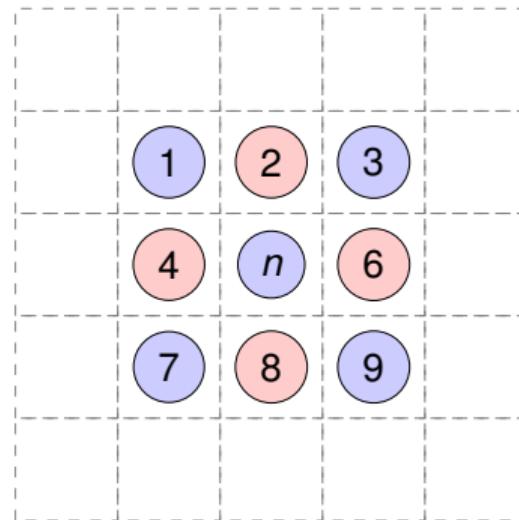


SAMPLING GRID AND NEIGHBORHOOD

- Let $n \in \Lambda$ be any point the grid.
- Let $\eta(n)$ be a neighborhood of $n \in \Lambda$.
- The first-order neighborhood consists of immediate top, bottom, left, and right neighbors of n .
- Let \mathcal{N} be a neighborhood system, a collection of neighborhoods of all $n \in \Lambda$.

$$n \notin \eta(n)$$

$$n \in \eta(l) \Leftrightarrow l \in \eta(n)$$



RANDOM FIELD

- A random field \mathcal{Y} over Λ is a multidimensional random process where each site $n \in \Lambda$ is assigned a random variable.
- A random field \mathcal{Y} with the following properties:

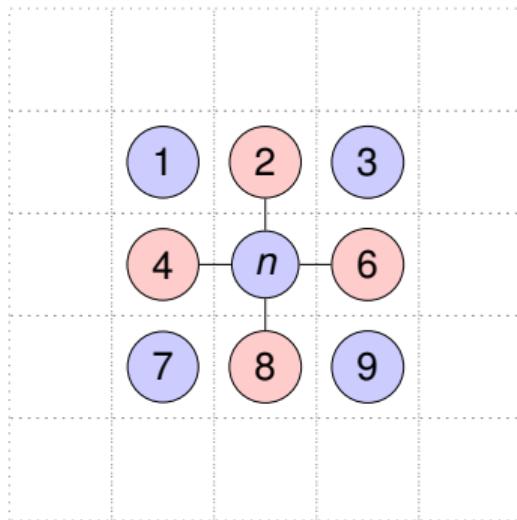
$$P(\mathcal{Y} = \nu) > 0 \quad \forall \nu \in \Gamma$$

$$P(\mathcal{Y}_n = \nu_n \mid \mathcal{Y}_I = \nu_I, \forall I \neq n) = P(\mathcal{Y}_n = \nu_n \mid \mathcal{Y}_I = \nu_I, \forall I \in \eta(n)) \quad \forall n \in \Lambda, \forall \nu \in \Gamma$$

P is a probability measure, and is called a Markov random field with state space Γ .

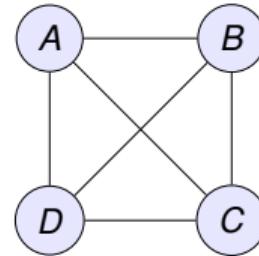
CLIQUE

- A clique c defined over Λ with respect to \mathcal{N} is a subset of Λ such that either c consists of a single site or every pair of sites in c are neighbors, that is, belong to η .
- The set of all cliques is denoted by C .
- A two-element spatial clique $\{n, l\}$ are two immediate horizontal, vertical or diagonal neighbors.



CLIQUE IN GENERAL

- Clique is a subset of vertices of an undirected graph such that every two distinct vertices in the clique are adjacent.
- A **clique** is a subset of nodes in which every node is connected to every other node.
- A **maximal clique** is a clique which cannot be extended by the addition of another node.



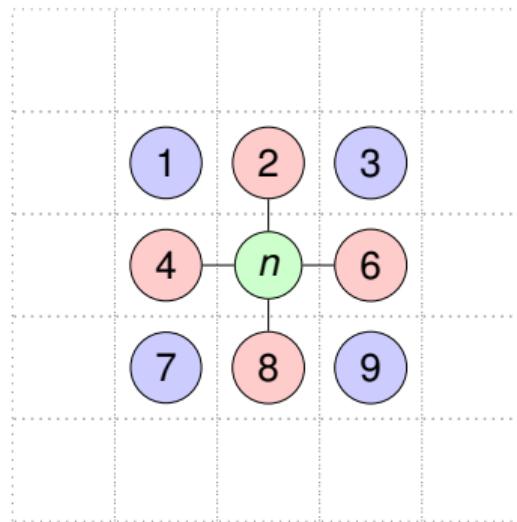
MARKOV RANDOM FIELD

- A Markov Random Field (MRF) is a graphical model of a joint probability distribution.
- Edges encode conditional independence.
- Let X_S be the set of random variables associated with the set of nodes S .
- Rule: Given disjoint subsets of nodes A , B and C , X_A is conditionally independent of X_B given X_C if there is no path from any node in A to any node in B that doesn't pass through a node of C . relationships.
- Markov property tells us that the joint distribution of X is determined entirely by the **local conditional distributions** $P(X_n \mid X_{\eta(n)})$

MARKOV RANDOM FIELD

Given the pink nodes, the green node is conditionally independent of all other nodes.

$$n \perp \{1, 3, 7, 9\} \mid \{2, 4, 6, 8\}$$



GIBBS DISTRIBUTION

- A Gibbs distribution on grid Λ and neighborhood \mathcal{N} takes the form:

$$P(\mathcal{Y} = \nu) = \frac{1}{Z} \exp\left(\frac{-1}{T} U(\nu)\right)$$

Z - Partition Function

T - temperature and is often taken to be 1.

U - energy function or potential function

$$U(\nu) = \sum_{c \in C} V(\nu, c)$$

For two element clique $\{n, l\}$ then $U(n, l) = V(\nu[n], \nu[l])$

MRF EQUIVALENCE TO GIBBS DISTRIBUTION

- The equivalence between Markov random fields and Gibbs distributions is provided through the important **Hammersley-Clifford theorem**.
- Theorem states that \mathcal{Y} is a MRF on Λ with respect to \mathcal{N} if and only if its probability distribution is a Gibbs distribution with respect to Λ and \mathcal{N} .

MAP ESTIMATION

- Let Y be a random field of observations
- Let \mathcal{Y} be a random field modeling the quantity we want to estimate based on Y
- .
- Let y, ν be their respective realizations.
- y could be a difference between two images.
- ν could be a field of motion detection labels.
- To compute ν based on y (MAP) estimation is

$$\hat{\nu} = \arg \max_{\nu} \underbrace{P(\mathcal{Y} = \nu \mid y)}_{\text{maximum of the posterior probability}} = \arg \max_{\nu} \underbrace{P(\mathcal{Y} = y \mid \nu) \cdot P(\mathcal{Y} = \nu)}_{\text{maximum likelihood}}$$

ML ESTIMATION

- To compute ν based on y (MAP) estimation is

$$\hat{\nu} = \arg \max_{\nu} \underbrace{P(\mathcal{Y} = \nu \mid y)}_{\text{maximum of the posterior probability}} = \arg \max_{\nu} \underbrace{P(\mathcal{Y} = y \mid \nu) \cdot P(\mathcal{Y} = \nu)}_{\text{maximum likelihood}}$$

- If $P(\mathcal{Y} = \nu)$ is the same for all realizations ν , then only the likelihood $P(\mathcal{Y} = y \mid \nu)$ is maximized, resulting in the maximum likelihood (ML) estimation.

TABLE OF CONTENTS

- 1 MODULE 2
- 2 INTRODUCTION TO MOTION
- 3 MATH PRELIMINARIES
- 4 MOTION DETECTION
- 5 MOTION ESTIMATION

GOAL OF MOTION DETECTION

- Identify which image points, or, regions of the image, have moved.
- Motion detection applies to images acquired with a **static camera**.
- Motion of image points is not perceived directly but rather through intensity changes.
- Such intensity changes over time may be also induced by camera noise or illumination variations.

HYPOTHESIS TESTING

- Simplest motion detection algorithms
- Let H_S and H_M be two hypotheses.
- H_S declaring an image point at n as stationary (S). State S means 0.
- H_M says an image point at n is moving (M). State M means 1.
- Let q be noise. Noise is assumed as zero-mean Gaussian with variance σ^2 in stationary areas and uniformly distributed in range $[-L, L]$ in moving areas.
- Assume $I_k[n] = I_{k-1}[n] + q$.
- P_S is assumed Gaussian, while P_M is assumed uniform.
- The motivation is that in stationary areas only camera noise will distinguish same-position pixels, whereas in moving areas this difference is attributed to motion and therefore unpredictable.

HYPOTHESIS TESTING WITH FIXED THRESHOLD

- Let an observation, upon which we intend to select one of the two hypotheses be

$$\rho_k[n] = I_k[n] - I_{k-1}[n]$$

- Hypothesis test is:

$$\rho_k^2[n] > \theta \quad \forall n \in M$$

$$\rho_k^2[n] < \theta \quad \forall n \in S$$

$$\theta = 2\sigma^2 \ln \left(\vartheta \cdot 2L \cdot P_S / (\sqrt{2\pi\sigma^2} \cdot P_M) \right)$$

- Not robust to noise in the image
- For small θ "noisy" detection masks result
- For large θ only object boundaries and its most textured parts are detected.

HYPOTHESIS TESTING, FIXED THRESHOLD, AVERAGING

- Averaging the observations over an N -point spatial window W_n centered at n

$$\frac{1}{N} \sum_{m \in W_n} \rho_k^2[m] > \theta \quad \forall n \in M$$

- Used to attenuate the impact of noise

MOTION DETECTION BASED ON FRAME DIFFERENCES

- Motion detection based on frame differences (last 3 slides) does not perform well for large, untextured objects (e.g., a large, uniformly colored truck).
- Only pixels n where $|I_k[n] - I_{k-1}[n]|$ is sufficiently large can be reliably detected.
- Such pixels concentrate in narrow areas close to moving boundaries where object intensity is distinct from the background in the previous frame.

MOTION DETECTION BY COMPARING BACKGROUND INTENSITY

- Comparing the current intensity $I_k[n]$ to background intensity $B_k[n]$ instead of the previous frame $I_{k-1}[n]$.

$$\rho_k[n] = I_k[n] - B_k[n]$$

- Estimate $B_k[n]$ by means of temporal averaging or median filtering the intensity at each n .
- Median can suppress intensities associated with moving objects (large window) and is fast. It fails in the presence of parasitic motion, such as fluttering leaves or waves on water surface.

HYPOTHESIS TESTING WITH FIXED THRESHOLD

Non-parametric distributions

- At each location n of frame k , an estimate of the stationary (background) probability distribution is computed from K recent frames as

$$P_S(I_k[n]) = \frac{1}{K} \sum_{i=1}^K \kappa(I_k[n] - I_{k-i}[n])$$

κ is a zero-mean Gaussian with variance σ^2 considered as constant

- Hypothesis test:

$$P_S(I_k[n]) > \theta \quad \forall n \in S$$

$$P_S(I_k[n]) < \theta \quad \forall n \in M$$

$$\theta = \frac{P_M}{2L\vartheta P_S}$$

MAP MRF FORMULATION

- To find a MAP estimate of the random field E_k , maximize the posterior probability

$$P(E_k = e_k \mid \rho_k)$$

- Let $|I_k[n] - I_{k-1}[n]|$ be an observation modelled as

$$\rho_k[n] = \xi(e_k[n]) + q[n]$$

where q is zero-mean uncorrelated Gaussian noise with variance σ^2

$$\xi(e_k[n]) = \begin{cases} 0 & \text{if } e_k[n] = \mathcal{S} \\ \alpha & \text{if } e_k[n] = \mathcal{M} \end{cases}$$

α is average temporal intensity difference based on previous-time moving labels e_{k-1}

MAP MRF FORMULATION

- ξ attempts to closely model the observations since for a static image point it is zero, whereas for a moving point it tracks average temporal intensity mismatch.
- Likelihood $P(R_k = \rho_k | e_k)$
- Gibbs distribution for the a priori probability $\pi(E_k = e_k)$
- The overall energy function

$$U(\rho_k, e_{k-1}, e_k) = \frac{1}{2\sigma^2} \sum_n ((\rho_k[n] - \xi(e_k[n]))^2 + \sum_{[n,l] \in C} V_s(e_k[n], e_k[l]) + \sum_{[t_{k-1}, t_k]} V_t(e_{k-1}[n], e_k[n]))$$

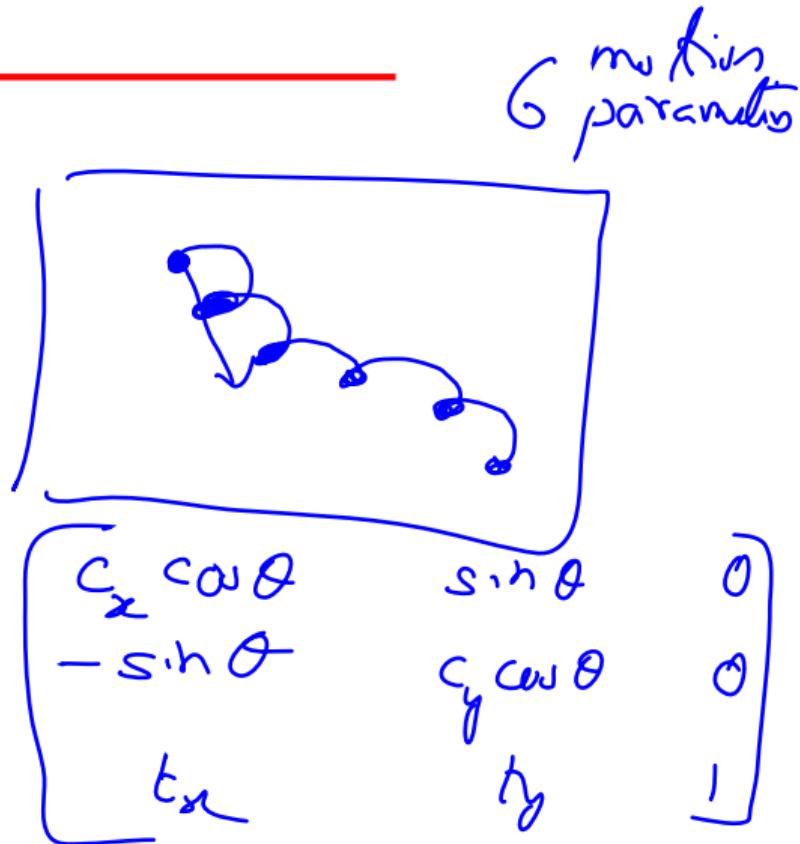
MRF FORMULATION

- The first term measures how well each label at n explains the observation $\rho_k[n]$.
- The other terms measure how contiguous the labels are in the image plane (V_s) and in time (V_t).
- A simple MRF model supported on the second-order neighborhood with two-element cliques $c = [n, l]$

$$V(e_k[n], e_k[l]) = \begin{cases} 0 & \text{if } e_k[n] = e_k[l] \\ \beta & \text{if } e_k[n] \neq e_k[l] \end{cases}$$

TABLE OF CONTENTS

- ① MODULE 2
- ② INTRODUCTION TO MOTION
- ③ MATH PRELIMINARIES
- ④ MOTION DETECTION
- ⑤ MOTION ESTIMATION



CONCEPT OF MOTION

- Video Compression
 - ▶ Reduce the number of bits needed to represent a video sequence.
 - ▶ Estimated motion parameters should lead to the highest compression ratio possible.
- Video Processing
 - ▶ Methods that improve quality
 - ▶ Eg: motion-compensated noise reduction, motion-compensated interpolation, and motion-based video segmentation.
 - ▶ True motion of image points is estimated.
 - ▶ Eg: In motion-compensated temporal interpolation, the task is to compute new images located between existing images of a video sequence). The new images should be consistent with the existing ones, image points belonging to moving objects must be displaced according to the true motion as otherwise "jerky" motion of objects would result.

MOTION ESTIMATION ALGORITHM

Three important elements

① Motion Models

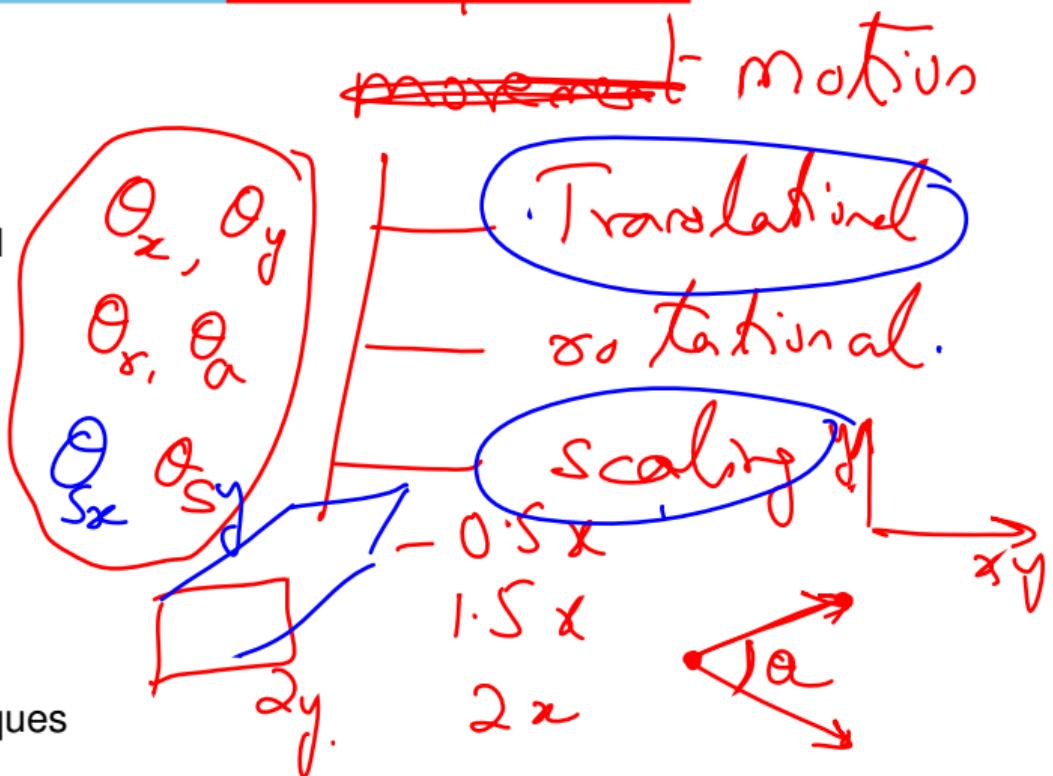
- ① Spatial Motion Model
- ② Temporal Motion Model
- ③ Region of Support
- ④ Observation Model

② Estimation criteria

- ① Pixel-Domain Criteria
- ② Regularization

③ Search Strategies

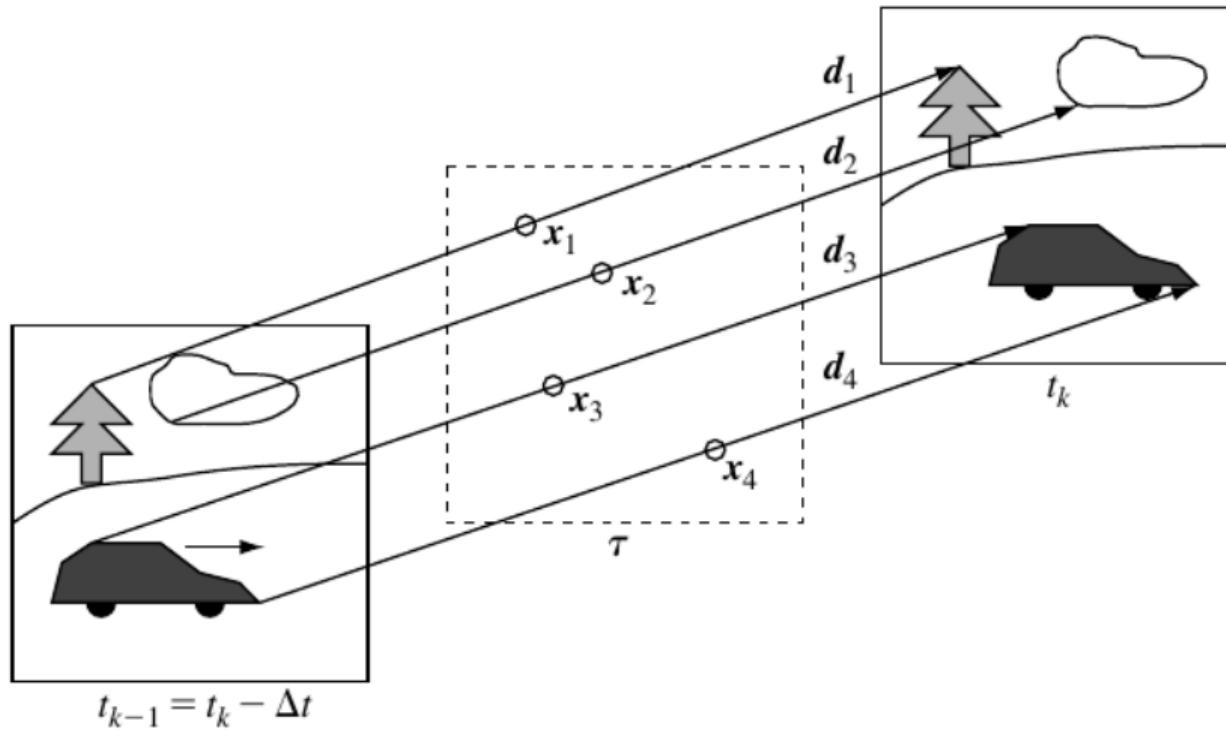
- ① Matching
- ② Gradient-based techniques



MOTION MODEL

- A motion model specifies how to represent motion in an image sequence.
- A model relating motion parameters to image intensities is called an observation model.

MOTION MODEL

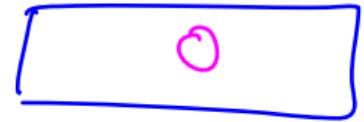


SPATIAL MOTION MODELS

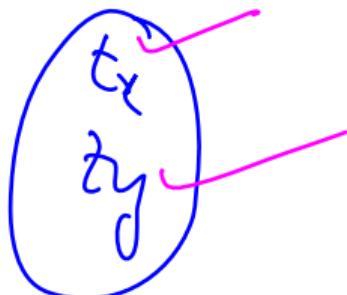
- The goal is to estimate the motion of image points.
- Object induced motion depends on the following:
 - ① image formation model for example, perspective, orthographic projection
 - ② motion model of 3D object, for example, rigid-body with 3D translation and rotation, 3D affine motion
 - ③ surface model of 3D object, for example, planar, parabolic.

Perspective, rigid-body, planes

SPATIAL MOTION MODEL – TRANSLATIONAL MODEL

 t  $t+1k$ 

- Orthographic projection and arbitrary 3D surface undergoing 3D translation



Velocity vector $v(\mathbf{x}) = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$

Parameters $b = (b_1, b_2)^\top$

- Relatively simple and extensively used in practice.

0.10s



0.30s
 $t + 10$



SPATIAL MOTION MODEL – PARAMETRIC MODEL

TRS

- Orthographic projection and 3D affine motion of a planar surface

Velocity vector $\nu(\mathbf{x}) = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} + \begin{pmatrix} b_3 & b_4 \\ b_5 & b_6 \end{pmatrix} \mathbf{x}$

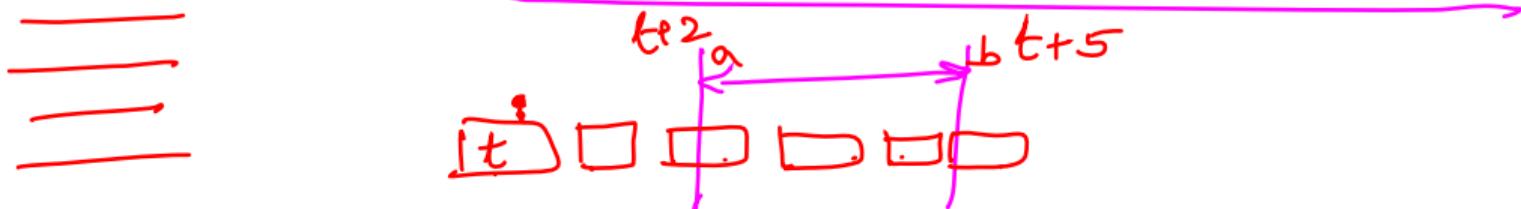
Parameters $b = (b_1, \dots, b_6)^\top$

- Simple and powerful
- A complex model applied to a small region of support may lead to an actual increase in the estimation error compared to a simpler model.

TEMPORAL MOTION MODELS



- The trajectories of individual image points drawn in the (x, y, t) space of an image sequence depend on object motion.
- Assume that the velocity $v_t(\mathbf{x})$ is constant between $t = t_{k-1}$ and $\tau (\tau > t)$
- $\mathbf{d}_{t,\tau}(\mathbf{x})$ is a displacement vector measured in the positive direction of time, from t to τ .
- The task is to find the two components of velocity or displacement at each \mathbf{x} .



TEMPORAL MOTION MODELS

t

$$\tau = t + k$$

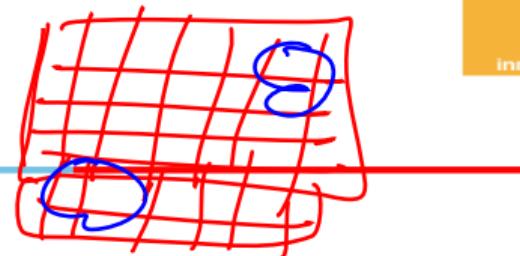
- Linear Trajectory (two velocity (linear) variables)

$$\begin{aligned}\mathbf{x}(\tau) &= \mathbf{x}(t) + \nu_t(\mathbf{x}) \cdot (\tau - t) \\ &= \mathbf{x}(t) + \mathbf{d}_{t,\tau}(\mathbf{x})\end{aligned}$$

- Quadratic Trajectory (two velocity (linear) variables and two acceleration (quadratic) variables $a = (a_1, a_2)^\top$)

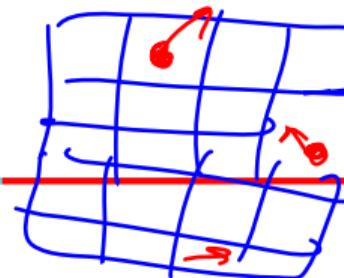
$$\mathbf{x}(\tau) = \mathbf{x}(t) + \nu_t(\mathbf{x}) \cdot (\tau - t) + \frac{1}{2} \cdot \mathbf{a}_t(\mathbf{x}) \cdot (\tau - t)^2$$

REGION OF SUPPORT



- The set of points \mathbf{x} to which spatial and temporal motion models apply is called the region of support \mathcal{R} .
- For a given motion model, the smaller the region of support , the better the approximation of motion.
- Four types
 - ① $\mathcal{R} = \text{the whole image}$
 - ★ A single motion model applies to all image points.
 - ★ Most constrained model
 - ★ Very few parameters can approximate the motion of all image points.

REGION OF SUPPORT



② \mathcal{R} = one pixel

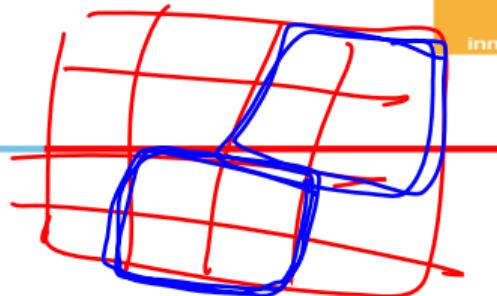
- ▶ This model applies to a single image point.
- ▶ Also called dense motion representation.
- ▶ Translational spatial model is used jointly with the linear temporal model.

$$\text{Velocity vector } \nu(\mathbf{x}) = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

$$\text{Linear Trajectory } \mathbf{x}(\tau) = \mathbf{x}(t) + \nu_t(\mathbf{x}) \cdot (\tau - t)$$

- ▶ Least constrained model
- ▶ A very large number of motion fields can be represented by all possible combinations of parameter values.
- ▶ High computational complexity

REGION OF SUPPORT



③ \mathcal{R} = rectangular block of pixels

- ▶ This motion model applies to a rectangular (or square) block of image points.
- ▶ Spatial translation model and temporal linear model of a square block of pixels is very powerful model and is used today in all digital video compression standards.
- ▶ Spatial translation and temporal quadratic motion is used in B frames of MPEG.

④ \mathcal{R} = irregularly-shaped region

- ▶ This model applies to all pixels in region of arbitrary shape.
- ▶ A square block divided into arbitrarily shaped parts, each with independent translational motion, is used in MPEG-4.

OBSERVATION MODELS

- Goal is to estimate motion based on intensity variations in time.
- Assume that objects do not change their appearance as they move. That is, image intensity $I = f(x, y, t)$ remains constant along motion trajectory s .

$$\frac{dI}{ds} = 0$$

$$\frac{\partial I}{\partial x} \nu_1 + \frac{\partial I}{\partial y} \nu_2 + \frac{\partial I}{\partial t} = 0 \quad \text{apply chain rule}$$

$$(\nabla I)^\top \nu + \frac{\partial I}{\partial t} = 0$$

- For I sampled in time, the constant-intensity assumption means that

$$I_{t_k}(x(t_k)) = I_{t_{k-1}}(x(t_{k-1})) \leftarrow$$

OBSERVATION MODELS

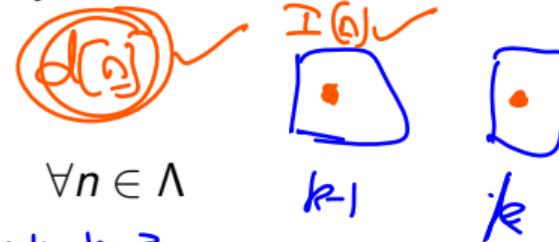
$$\frac{n^k I}{k-1} + d[n]$$

- Assume spatial sampling of intensities, linear trajectory, $t = t_{k-1}$ and $\tau = t_k$

$$\frac{n^k P}{k-1}$$

$$\mathbf{x}(\tau) = \mathbf{x}(t) + \nu_t(\mathbf{x}) \cdot (\tau - t)$$

$$\frac{dI}{ds} = I_k[n] - I_{k-1}[n - d[n]] = 0, \quad \forall n \in \Lambda$$



- The above equation does not hold exactly due to noise, aliasing, illumination variations, etc., and a minimization of some function of the above equation is needed. When scene illumination changes, a constraint based on the spatial gradient's constancy in the direction of motion can be used.
- In areas of uniform intensity but substantial color detail, the inclusion of a color-based constraint could prove beneficial. A multicomponent (vector) function replaces I .

MOTION ESTIMATION ALGORITHM

Three important elements

① Motion Models

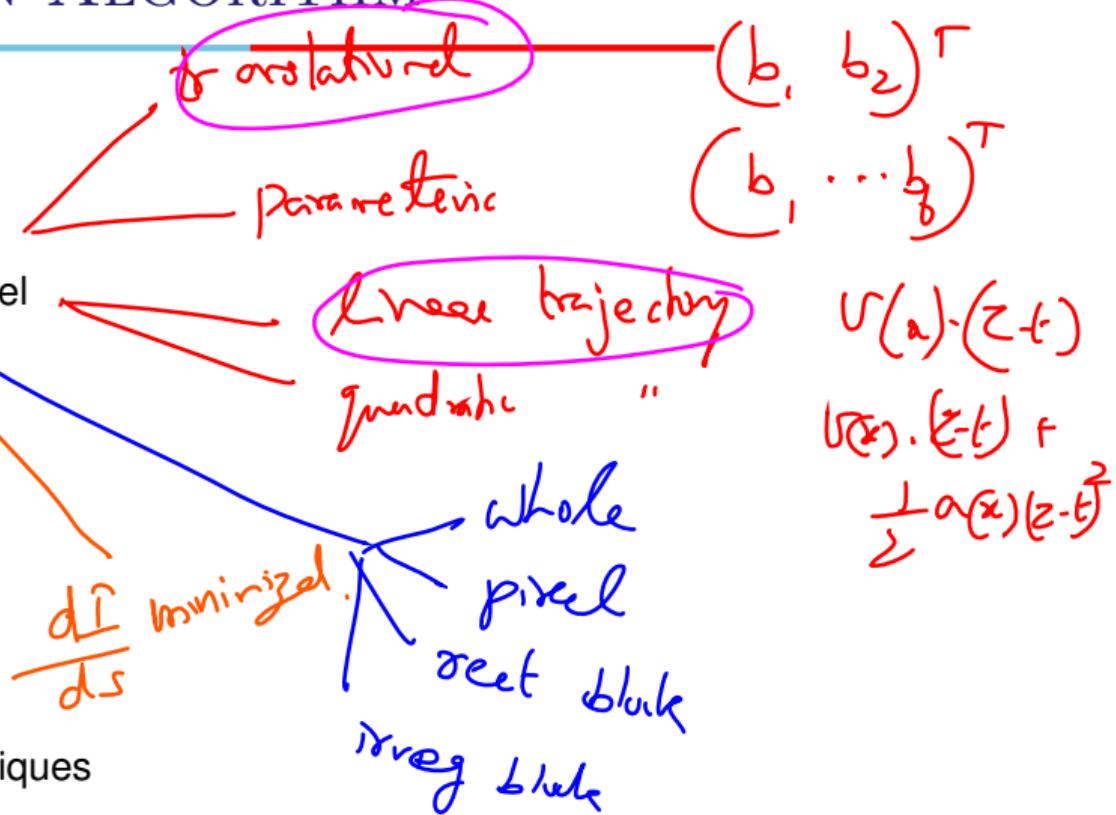
- ① Spatial Motion Model
- ② Temporal Motion Model
- ③ Region of Support
- ④ Observation Model

② Estimation criteria

- ① Pixel-Domain Criteria
- ② Regularization

③ Search Strategies

- ① Matching
- ② Gradient-based techniques



ESTIMATION CRITERIA

- Motion models are incorporated into an estimation criterion that will be optimized.
- There is no unique criterion for motion estimation because its choice depends on the task at hand.
- In compression an average performance or prediction error of a motion estimator is the criteria.
- In motion-compensated interpolation the worst case performance (maximum interpolation error) is the criteria.
- The selection of a criterion may be guided by the processor capabilities on which the motion estimation will be implemented.

PIXEL-DOMAIN CRITERIA

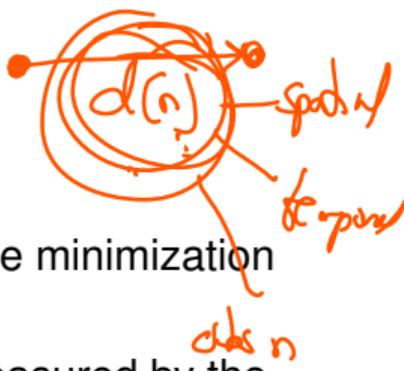
$$I_k(n) - I_{k-1}(n - d(n)) = 0$$

- Motion-compensated prediction of $I_k[n]$ is given by

$$\tilde{I}_k[n] = I_{k-1}[n - d[n]]$$

- Then Error is given as

$$\epsilon_k[n] = I_k[n] - \tilde{I}_k[n] \quad \forall n \in \Lambda$$



- Discrete version of the constant-intensity assumption aim at the minimization of a function $\epsilon_k[n]$.
- Similarity between $I_k[n]$ and its prediction $\tilde{I}_k[n]$ can be also measured by the following cross-correlation function (which needs to be maximised):

$$C(d) = \sum_n I_k[n]I_{k-1}[n - d[n]]$$



PIXEL-DOMAIN CRITERIA

- Estimation criterion is then

$$\mathcal{E}(d) = \sum_{n \in R} \Phi(I_k[n] - \tilde{I}_k[n])$$

- Φ is a non-negative real-valued function.
- Quadratic function – a single large error ϵ (an outlier) over-contributes to \mathcal{E} and biases the estimate of \mathbf{d} .

$$\Phi(\epsilon) = \epsilon^2$$

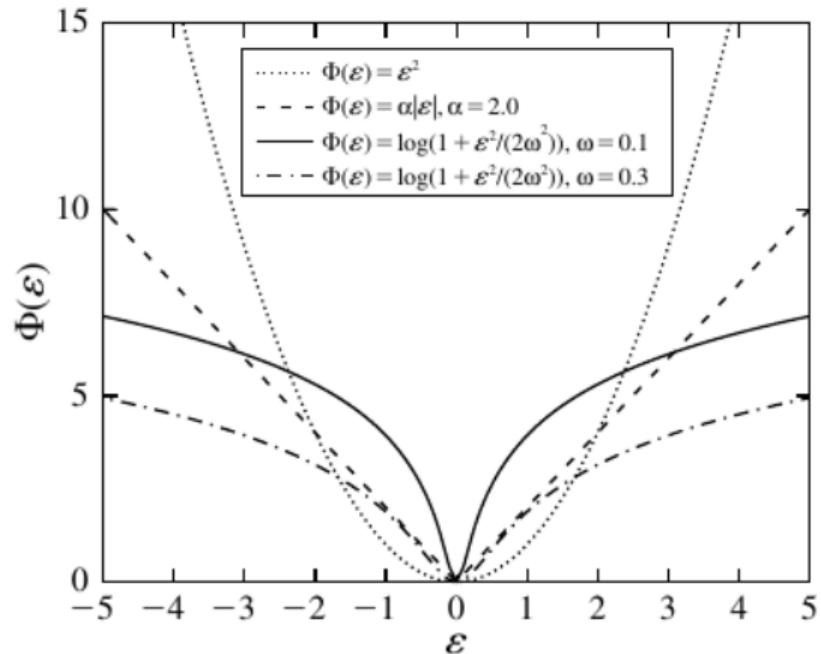
- Linear function – better choice, used in video encoders

$$\Phi(\epsilon) = \alpha(\epsilon)$$

- Lorentzian function – grows slower than $|x|$ for large errors.

$$\Phi(\epsilon) = \log(1 + \epsilon^2/2\omega^2)$$

PIXEL-DOMAIN ESTIMATION CRITERIA



REGULARIZATION

- Moving objects are close to being rigid.
- Motion field ν_t is locally smooth.
- Gradient is a good measure of local smoothness.
- Minimizing the following criterion : where D is the domain of the image. This formulation is often referred to as regularization.

$$E(\nu) = \int_D \left(\nabla^\top I(x)\nu(x) + \frac{\partial I(x)}{\partial t} \right)^2 + \lambda (\|\nabla(\nu_1(x))\|^2 + \|\nabla(\nu_2(x))\|^2) dx$$

MOTION ESTIMATION ALGORITHM

Three important elements

① Motion Models

- ① Spatial Motion Model
- ② Temporal Motion Model
- ③ Region of Support
- ④ Observation Model

② Estimation criteria

- ① Pixel-Domain Criteria
- ② Regularization

③ Search Strategies

- ① Matching
- ② Gradient-based techniques

SEARCH STRATEGIES

- Develop an efficient (complexity) and effective (solution quality) strategy for finding an estimate of motion parameters.

MATCHING

- For a small number of motion parameters
- A small state space for each of them
- Minimizing a prediction error
- Motion-compensated predictions $\tilde{I}_k[n]$ for various motion candidates \mathbf{d} are matched with $I_k[n]$ within the region of support of the motion model.
- The candidate yielding the best match for a given criterion becomes the optimal estimate.

GRADIENT-BASED TECHNIQUE

- Estimation criteria E is differentiable.
- To avoid non-linear optimization I is usually linearized using Taylor expansion with respect to $d[n]$.
- Gradient-based estimation yields accurate results only in regions of small motion.
- The approach fails if motion is large. This deficiency is usually compensated for by a hierarchical or multiresolution implementation.

GLOBAL MOTION ESTIMATION ALGORITHM



























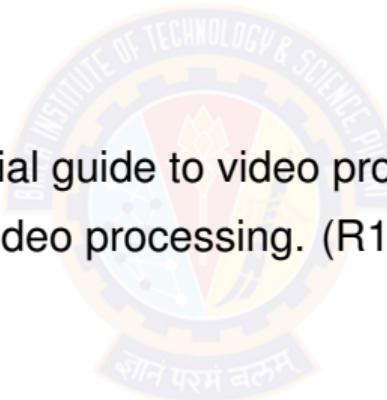






REFERENCES

- ① Bovik, Alan C. The essential guide to video processing. (T1) Ch 3
- ② Tekalp, A. Murat. Digital video processing. (R1)



Thank You!



VIDEO ANALYTICS MODULE # 3 : VIDEO ENHANCEMENT AND RESTORATION



BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Seetha Parameswaran
BITS Pilani

The instructor is gratefully acknowledging
the authors who made their course
materials freely available online.

This deck is prepared by Seetha Parameswaran.

TABLE OF CONTENTS

- ① MODULE 3 TOPICS
- ② VIDEO ENHANCEMENT AND RESTORATION
- ③ FILTERING PRELIMS
- ④ SPATIOTEMPORAL NOISE FILTERING

MODULE TOPICS....

- Spatio temporal noise filtering
- Coding Artifact reduction
- Blotch reduction and removal
- Vinegar Syndrome removal
- Kinescope moiré removal
- Flicker correction
- Scratch removal
- Application

TABLE OF CONTENTS

- ① MODULE 3 TOPICS
- ② VIDEO ENHANCEMENT AND RESTORATION
- ③ FILTERING PRELIMS
- ④ SPATIOTEMPORAL NOISE FILTERING

VIDEO ENHANCEMENT – ITS NEED

- Video or recorded image sequences suffer from severe degradations.
- Due to
 - ▶ imperfect or uncontrollable recording conditions, such as one encounters in astronomy, forensic sciences, and medical imaging.
 - ▶ visible coding artifacts, such as blocking, ringing, and mosquito noise.
- Improve the visual quality
- Increase the performance of subsequent tasks such as analysis and interpretation.

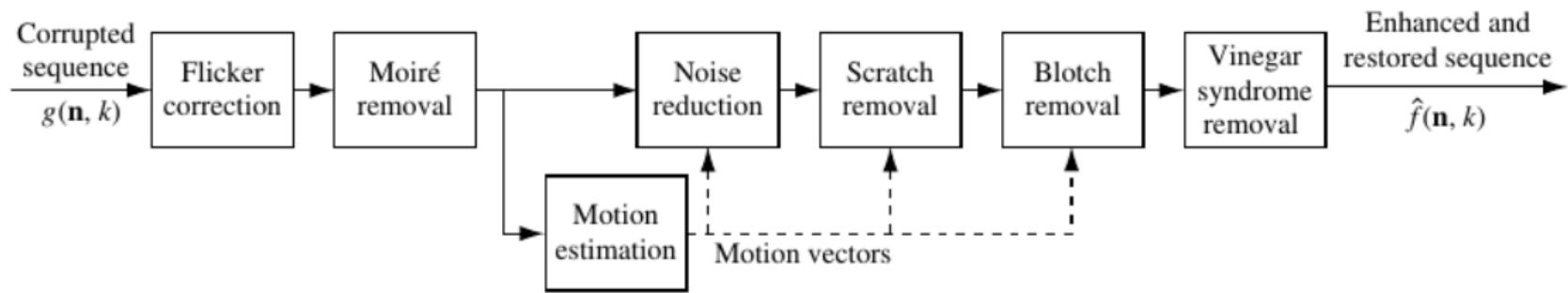
VIDEO RESTORATION – ITS NEED

- Restoration is preserving motion pictures and video tapes recorded over the last century.
- The records are deteriorating rapidly due to aging effects of the physical reels of film and magnetic tapes that carry the information.
- Restoration process
 - ① Video is transferred from the original film reels or magnetic tape to digital media.
 - ② Then, all kinds of degradations are removed from the digitized image sequences.
- Objective of restoration
 - ▶ remove irrelevant information such as noise and blotches
 - ▶ restores the original spatial and temporal correlation structure of digital image sequences.

CHALLENGE

- Large amount of data in a video
- Enhancement and Restoration methods for image sequences should have a manageable complexity and should be semiautomatic.
 - ▶ Semiautomatic indicates that professional operators control the visual quality of the restored image sequences by selecting values for some of the critical restoration parameters.

STEPS IN VIDEO ENHANCEMENT AND RESTORATION



Video enhancement and restoration techniques are sometimes referred to as **spatiotemporal filters or 3D filters**.

STEPS IN VIDEO ENHANCEMENT AND RESTORATION

NOISE REMOVAL spatial and temporal noise to be removed

CODING ARTIFACT REDUCTION OR BLOCKINESS REDUCTION due to the lossy Discrete Cosine Transform

BLOTTCHES are dark and bright spots that are often visible in damaged film image sequences. The removal of blotches is a temporal detection and interpolation problem.

VINEGAR SYNDROME represents a special type of impairment related to film (e.g., partial loss of color, blur).

STEPS IN VIDEO ENHANCEMENT AND RESTORATION

INTENSITY FLICKER refers to variations in intensity in time, caused by aging of film, by copying and format conversion (e.g., from film to video), or by variations in shutter time.

KINESCOPE MOIRÉ phenomenon appears during film-to-video transfer using telecine devices.

FILM SCRATCHES are either bright or dark vertical lines spanning the entire frame. They appear approximately at the same place in consecutive frames.

TABLE OF CONTENTS

- 1 MODULE 3 TOPICS
- 2 VIDEO ENHANCEMENT AND RESTORATION
- 3 FILTERING PRELIMS
- 4 SPATIOTEMPORAL NOISE FILTERING

FILTERS

- Filtering refers to accepting (passing) or rejecting certain components.
- Filters are also called spatial masks, kernels, templates, and windows.
- A **spatial filter** consists of
 - ① a neighborhood, (typically a small rectangle)
 - ② a predefined operation that is performed on the image pixels encompassed by the neighborhood.
- Filtering creates a new pixel with coordinates equal to the coordinates of the center of the neighborhood, and whose value is the result of the filtering operation.
- A processed (filtered) image is generated as the center of the filter visits each pixel in the input image. If the operation performed on the image pixels is linear, then the filter is called a **linear spatial filter**.

LINEAR SPATIAL FILTER

Pixels of Image section under filter

$f(x - 1, y - 1)$	$f(x - 1, y)$	$f(x - 1, y + 1)$
$f(x, y - 1)$	$f(x, y)$	$f(x, y + 1)$
$f(x + 1, y - 1)$	$f(x + 1, y)$	$f(x + 1, y + 1)$

Filter coefficients

$w(-1, -1)$	$w(-1, 0)$	$w(-1, 1)$
$w(0, -1)$	$w(0, 0)$	$w(0, 1)$
$w(1, -1)$	$w(1, 0)$	$w(1, 1)$

$$\begin{aligned}
 g(x, y) = & w(-1, -1)f(x - 1, y - 1) + w(-1, 0)f(x - 1, y) + w(-1, 1)f(x - 1, y + 1) \\
 & + w(0, -1)f(x, y - 1) + w(0, 0)f(x, y) + w(0, 1)f(x, y + 1) \\
 & + w(1, -1)f(x + 1, y - 1) + w(1, 0)f(x + 1, y) + w(1, 1)f(x + 1, y + 1)
 \end{aligned}$$

Coefficient of the filter, $w(0, 0)$ aligns with the pixel at location (x, y) .

CORRELATION

- Correlation of a filter $w(x, y)$ of size $m \times n$ with an image $f(x, y)$
- Denoted as $w(x, y) \otimes f(x, y)$

$$w(x, y) \otimes f(x, y) = \sum_{s=-a}^a \sum_{t=-b}^b w(s, t) f(x + s, y + t)$$

$$a = (m - 1)/2, b = (n - 1)/2$$

- This equation is evaluated for all values of the displacement variables x and y so that all elements of w visit every pixel in f , where we assume that f has been padded appropriately.

CORRELATION

		Padded f		
		0 0 0 0 0 0 0 0 0 0		
		0 0 0 0 0 0 0 0 0 0		
		0 0 0 0 0 0 0 0 0 0		
Origin $f(x, y)$		0 0 0 0 0 0 0 0 0 0		
0 0 0 0 0		0 0 0 0 0 1 0 0 0 0		
0 0 0 0 0		0 0 0 0 0 0 0 0 0 0		
$w(x, y)$		0 0 0 0 0 0 0 0 0 0		
0 0 1 0 0		0 0 0 0 0 0 0 0 0 0		
1 2 3		0 0 0 0 0 0 0 0 0 0		
0 0 0 0 0		0 0 0 0 0 0 0 0 0 0		
4 5 6		0 0 0 0 0 0 0 0 0 0		
0 0 0 0 0		0 0 0 0 0 0 0 0 0 0		
7 8 9		0 0 0 0 0 0 0 0 0 0		
(a)		(b)		
Initial position for w		Full correlation result	Cropped correlation result	
1 2 3 0 0 0 0 0 0 0		0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0	
4 5 6 0 0 0 0 0 0 0		0 0 0 0 0 0 0 0 0 0	0 9 8 7 0	
7 8 9 0 0 0 0 0 0 0		0 0 0 0 0 0 0 0 0 0	0 6 5 4 0	
0 0 0 0 0 0 0 0 0 0		0 0 0 9 8 7 0 0 0 0	0 3 2 1 0	
0 0 0 0 1 0 0 0 0 0		0 0 0 6 5 4 0 0 0 0	0 0 0 0 0 0 0 0 0 0	
0 0 0 0 0 0 0 0 0 0		0 0 0 3 2 1 0 0 0 0		
0 0 0 0 0 0 0 0 0 0		0 0 0 0 0 0 0 0 0 0		
0 0 0 0 0 0 0 0 0 0		0 0 0 0 0 0 0 0 0 0		
0 0 0 0 0 0 0 0 0 0		0 0 0 0 0 0 0 0 0 0		
(c)		(d)	(e)	

CORRELATION

- A function that contains a single 1 with the rest being 0s a **discrete unit impulse**.
- Correlation of a function with a discrete unit impulse yields a rotated version of the function at the location of the impulse.
- Correlation can be used to find matches between images.

CONVOLUTION

- Convolution of a filter $w(x, y)$ of size $m \times n$ with an image $f(x, y)$
- Denoted as $w(x, y) \oplus f(x, y)$

$$w(x, y) \oplus f(x, y) = \sum_{s=-a}^a \sum_{t=-b}^b w(s, t) f(x - s, y - t)$$

$$a = (m - 1)/2, b = (n - 1)/2$$

- This equation is evaluated for all values of the displacement variables x and y so that all elements of w visit every pixel in f , where we assume that f has been padded appropriately.

CONVOLUTION

Rotated w	Full convolution result	Cropped convolution result
9 8 7 6 5 4 3 2 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 1 2 3 0 0 0 0 0 0 0 4 5 6 0 0 0 0 0 0 0 7 8 9 0	0 0 0 0 0 0 0 1 2 3 0 0 4 5 6 0 0 7 8 9 0
(f)	(g)	(h)

CONVOLUTION

- Convolution of a function with an impulse copies the function at the location of the impulse.
- If the filter mask is symmetric, correlation and convolution yield the same result.
- **Convolving a mask with an image** often is used to denote the sliding, sum-of-products process, and does not necessarily differentiate between correlation and convolution.

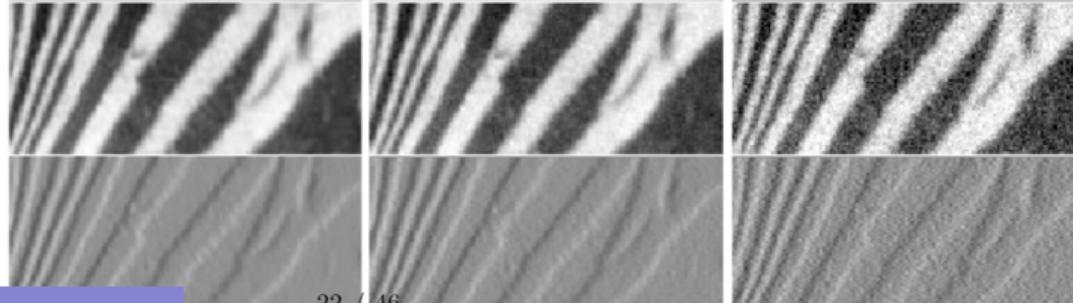
TABLE OF CONTENTS

- 1 MODULE 3 TOPICS
- 2 VIDEO ENHANCEMENT AND RESTORATION
- 3 FILTERING PRELIMS
- 4 SPATIOTEMPORAL NOISE FILTERING

NOISE

- Noise is anything in the image that we are not interested in.
 - ▶ Light fluctuations
 - ▶ Sensor noise or Camera noise
 - ▶ Quantization effects
 - ▶ Thermal noise
 - ▶ Granular noise on film
 - ▶ Shot noise originating in electronic hardware and storage on magnetic tape
- Effects of the noise are nonlinear of nature.

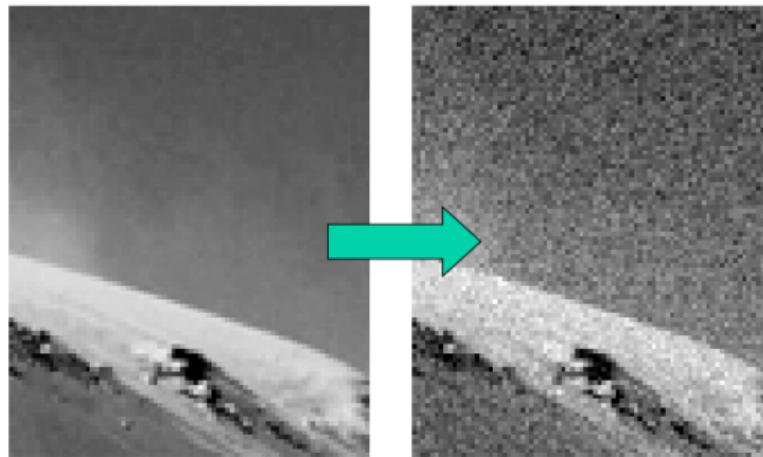
Increasing noise



GAUSSIAN NOISE

- Also called White noise.
- Generated by the Gaussian curve.

mean 0, sigma = 16



NOISE REMOVAL

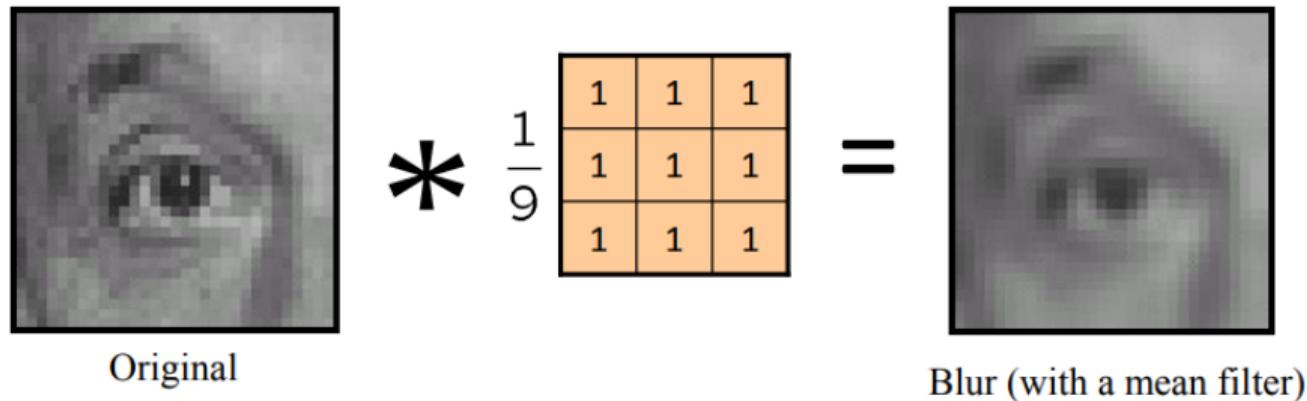
- The aggregated effect of noise is modeled as an additive white (sometimes Gaussian) process with zero mean and variance σ^2 that is independent of the ideal uncorrupted image sequence $f(n_1, n_2, k)$.
- The recorded image sequence

$$g(n_1, n_2, k) = f(n_1, n_2, k) + w(n_1, n_2, k)$$

- The objective of noise reduction is to make an estimate $\hat{f}(n, k)$ of the original image sequence given only the observed noisy image sequence $g(n_1, n_2, k)$.

SMOOTHING FILTER

- Smoothing reduces noise.



- The filter is called averaging filter or mean filter or box filter.
- Result is blurred image.

SMOOTHING + SHARPENING FILTER

- Sharpen the smoothed image to reduce the blur and noise.



$$\text{Original} \quad * \left(\begin{matrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{matrix} - \frac{1}{9} \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix} \right) = \text{Sharpening filter output}$$



Sharpening filter
(accentuates edges)

GAUSSIAN SMOOTHING FILTER

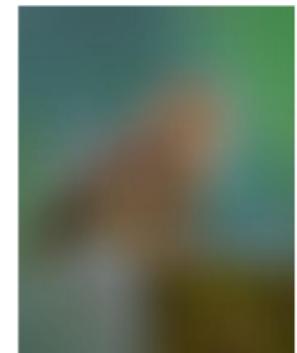
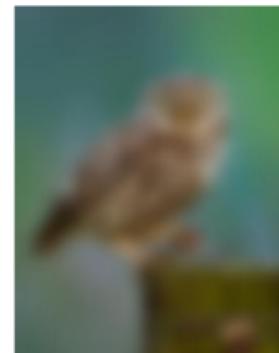
- The coefficients are a 2D Gaussian.
- Gives more weight at the central pixels and less weights to the neighbours.
The farther away the neighbours, the smaller the weight.

$$G = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)$$

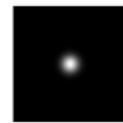
$$w = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

$$w = \frac{1}{273} \begin{bmatrix} 1 & 4 & 7 & 4 & 1 \\ 4 & 16 & 26 & 16 & 4 \\ 7 & 26 & 41 & 26 & 7 \\ 4 & 16 & 26 & 16 & 4 \\ 1 & 4 & 7 & 4 & 1 \end{bmatrix}$$

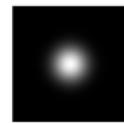
GAUSSIAN SMOOTHING FILTER



$\sigma = 1$ pixel



$\sigma = 5$ pixels



$\sigma = 10$ pixels



$\sigma = 30$ pixels

SEPERABLE GAUSSIAN SMOOTHING FILTER

$$G = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) \exp\left(-\frac{y^2}{2\sigma^2}\right) = g(x)g(y)$$

$$w = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} [1 \quad 2 \quad 1]$$

$$w = \frac{1}{256} \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 6 \\ 4 \\ 1 \end{bmatrix} [1 \quad 4 \quad 6 \quad 4 \quad 1]$$

GAUSSIAN SMOOTHING FILTER

Apply the Gaussian filter to the image:
 Borders: keep border values as they are

15	20	25	25	15	10
20	15	50	30	20	15
20	50	55	60	30	20
20	15	65	30	15	30
15	20	30	20	25	30
20	25	15	20	10	15

Original image

1	2	1
2	4	2
1	2	1

Or:

$$\frac{1}{4} \begin{bmatrix} 1 & 2 & 1 \end{bmatrix}$$

$$\frac{1}{4} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

$$* \frac{1}{16}$$

15	20	24	23	16	10
20	25	36	33	21	15
20	44	55	51	35	20
20	29	44	35	22	30
15	21	25	24	25	30
20	21	19	16	14	15
15	20	24	23	16	10
19	28	38	35	23	15
20	35	48	43	28	21
19	31	42	36	26	28
18	23	28	25	22	21
20	21	19	16	14	15

27

Birla Institute of Technology and Science, Pilani
Work Integrated Learning Programmes Division
M. Tech. in AI & ML
Mid-Semester Sample QP

Course Number AIMLCZG531
 Course Name VIDEO ANALYTICS
 Nature of Exam Closed Book

1. What is the difference between frame rate, refresh rate and frame size? [3]
2. What are the formats for composite analog video. [3]

3. Find the file size of a video size that has 1 hour and 45 minutes, 24-bit color encoding per pixel and has 1920x1080 resolution. [5]

$$\text{Total Pixels} = 1920 * 1080 = 2073600$$

$$\text{Size of Each Frame} = \text{Total Pixels} * 24 \text{ bit} = 2073600 * 24 = 49766400 \text{ bits}$$

$$\text{Video Length} = 1 \text{ hour and 45 minutes} = 105 \text{ minutes} = 6300 \text{ seconds}$$

$$\text{Video Size} = \text{Frame rate} * \text{Size of Each frame} * \text{Video Length}$$

$$= 24 * 49766400 * 6300 = 7524679680000 \text{ bits} = 7524679680000 / 8 \text{ bytes} = 940584960000 \text{ Bytes} = 940584960000 / 2^{30} \text{ GB} = 875.988 \text{ GB}$$

$$(1 \text{ GByte} = 1024 \text{ MByte} = 1024 * 1024 * 1024 \text{ Bytes} = 2^{30} \text{ Bytes}).$$

4. What is video scanning and what are the two types of video scanning. Explain. [5]
5. What are the three types of frames see in MPEG-4? Explain. [5]
6. Consider the frames given below at time $k - 1$ and k . Compute is the motion vector using frame differences. [3]

Frame I_{k-1}			
3	3	3	3
3	3	3	255
3	3	255	3
3	255	3	3

Frame I_k			
3	3	3	3
3	3	255	3
3	3	255	3
3	3	255	3

$$|I_k - I_{k-1}| = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 252 & 252 \\ 0 & 0 & 0 & 0 \\ 0 & 252 & 252 & 0 \end{bmatrix}$$

$$\forall n = 0 \in S$$

$$\forall n = 252, \in M$$

7. What is the difference between motion model and observation model? [2]
8. What are the two spatial models for motion estimation? [2]

9. What are the two temporal models for motion estimation? [2]
10. What are the four types of region of support used for motion estimation? [4]
11. Explain the characteristics of Digital Video.
12. Explain the functionalities in Video Analytics.
13. Discuss the Analog Video Formats with an example.
14. Explain Aspect Ratio in Digital Video. Compute the Aspect Ratio of a Standard TV and HD TV.
15. What are sampling and quantization in the context of video analytics, and how do they influence the representation and processing of video signals?
16. Explain the principles of sampling and quantization, their relationship, and the impact on video quality, resolution, and fidelity.
17. What are video detection and estimation, and how do they differ from traditional video processing techniques?
18. Explain the importance of video detection and estimation in applications such as surveillance, security, traffic monitoring, and human-computer interaction.
19. Describe common techniques and algorithms used for video detection and estimation, such as object detection, motion estimation, tracking, activity recognition, and anomaly detection.
20. Discuss the advantages, limitations, and computational requirements of different algorithms in various applications.
21. Identify specific applications or scenarios where video detection and estimation are critical, such as crowd analysis in public spaces, vehicle tracking on highways, facial recognition in security systems, or gesture recognition in interactive environments.
22. Analyze the challenges posed by complex scenes, occlusions, lighting variations, and other factors in real-world applications.
23. What is Gaussian noise, and how does it affect the quality of digital images and videos?
24. Explain the mathematical representation of Gaussian noise and its characteristics.
25. Describe how the presence of Gaussian noise affects image quality in terms of visibility, clarity, and overall visual perception.
26. Discuss the impact of Gaussian noise on image processing tasks such as image enhancement, restoration, and analysis.
27. Identify specific applications or scenarios where Gaussian noise is commonly encountered, such as medical imaging, satellite imaging, or digital photography.
28. Discuss the challenges posed by Gaussian noise in these applications and the importance of noise reduction techniques.