ANNAMACHARYA INSTITUTE OF TECHNOLOGY & SCIENCES
RAJAMPET
(Autonomous)

Department of ARTIFICIAL INTELLIGENCE AND DATA SCIENCE
Lecture Notes

Name of the Faculty: B.SIREESHA
Class: III B.Tech II Sem.
Branch: AI&DS
Name of the Course: COMPUTER VISION
Subject Code: 23A306FT
Academic Year: 2025-2026

UNIT–I:LINEARFILTERS

IntroductiontoComputerVision

ComputerVisionisthefieldthatenablescomputersto**see,analyze,andunderstand images and videos** similar to human vision.

Computervisionisaninterdisciplinaryfieldofartificial intelligencethatenablesmachinesto interpret and gain high-level understanding from digital images or videos, essentially mimickinghumansight.

- **CoreConcepts:**Thefieldinvolvestaskssuchasobjectdetection(locatingobjects), objectrecognition(identifyingobjects),imageclassification(categorizingimages), and scene understanding (interpreting relationships between objects).
- **Process:** The typical workflow involves image acquisition, pre-processing (noise reduction,contrastenhancement),featureextraction(identifyingkeyelementslike edges, shapes), pattern recognition (comparing features to known examples), and decision-making.
- **Applications:**Itisusedacrossmanyindustriesincludinghealthcare(medicalimage analysis), autonomous vehicles (navigation, obstacle avoidance), manufacturing (quality control), and security (surveillance, face recognition).

GoalsofComputerVision

- Extractusefulinformationfromimages
- Detectobjects,edges,textures
- Recognizepatternsandshapes
- Understandscenestructure

Applications

- Facerecognition
- Medicalimageanalysis
- Autonomousvehicles
- Surveillancesystems
- Industrialinspection

DigitalImageRepresentation

A**digitalimage**is representedasa**2Dfunction**:

f(x,y)f(x,y)f(x,y)

Where:

- x,yx,yx,y→spatialcoordinates
- f(x,y)f(x,y)f(x,y)→intensity(grayvalue)

GrayLevelImage

- Valuesrangefrom**0to255**
- 0→Black
- 255→White

ColorImage

- Representedusing**RGB**
- Eachpixelhas3values:(R,G,B)

LinearFilters

A**linearfilter**modifiesanimageusinga**linearoperation**.

Linearfiltersaretransformationsappliedtoanimagetomodifysignals,oftenfornoise reduction, sharpening, or blurring.

PropertiesofLinearSystems

1. **Additivity**

T(f1+f2)=T(f1)+T(f2)T(f_1+f_2)=T(f_1)+T(f_2)T(f1+f2)=T(f1)+T(f2)

2. **Homogeneity**

T(af)=aT(f)T(af)= aT(f)T(af)=aT(f)

Ifbothpropertiesaresatisfied→systemis**linear**. **Convolution in**

Computer Vision

Convolutionisthe**coreoperation**inimagefiltering.

- **Convolution:**Thisisamathematicaloperationcentraltolinearfiltering.Itcombines aninputimagewithasmallermatrix ofweights,calleda **kernel**ormask,toproduce

anewimage.Theoutputpixel'svalueisaweightedsumofitsneighboringpixelsin the original image.

- **Properties:**Linearfiltersarecharacterizedbythepropertiesofsuperpositionand shift-invariance (moving the input shifts the output by the same amount).
- **Types:**Commontypesincludelow-passfilters(forsmoothing/blurring)andhigh-pass filters (for sharpening/edge detection).

MathematicalDefinition

g(x,y)=∑m∑nf(m,n)·h(x−m,y−n)g(x,y)=\sum_m\sum_nf(m,n)\cdoth(x-m,y-n)g(x,y)=m∑n∑ f(m,n)·h(x−m,y−n)

Where:

- fff→inputimage
- hhh→kernel(filtermask)
- ggg→outputimage

Kernel

- Smallmatrix(e.g.,3×3,5×5)
- Slidesovertheimage
- Determinestypeoffiltering

Example:AveragingFilter

19[111111111]\frac{1}{9}\begin{bmatrix}1& 1& 1\\1& 1&1\\1&1&1\end{bmatrix}91 111111111

Purpose→**Smoothing/Noiseremoval**

ShiftInvariantLinearSystems

Asystemis**shiftinvariant**ifshiftingtheinputimagecausestheoutputtoshiftbythesame amount.

A system is shift-invariant if a shift in the input signalcauses an identicalshift in the output signal, without changing the shape of the output. Linear, shift-invariant (LSI) systems are fundamentalincomputervisionandcanbeentirelycharacterizedbytheirimpulseresponse (the kernel).

T[f(x−x0,y−y0)]=g(x−x0,y−y0)T[f(x-x_0,y-y_0)]=g(x-x_0,y-y_0)T[f(x−x0,y−y0)]=g(x−x0,y−y0)

Importance

- Convolutionassumesshiftinvariance
- Mostimagefilterssatisfythisproperty

SpatialFrequency

Spatialfrequencyreferstohowrapidlyintensityvalueschangeinanimage.

- **SpatialFrequency:**Thisdescribeshowquicklypixelvalueschangeacrossanimage. Lowfrequencies representsmooth,gradualchanges (like largeshapes),whilehigh frequencies capture rapid changes (like edges and fine details).

FrequencyType	**Meaning**
Lowfrequency	Smoothregions
Highfrequency	Edges,noise

Fourier Transform(FT)

FourierTransformconvertsanimagefrom**spatialdomain→frequencydomain**.

- **FourierTransform(FT):** TheFTisacrucialtoolthatdecomposesanimagefromits originalspatialdomainintothefrequencydomain,representingitasasumofsine and cosine waves of different frequencies.
- **ConvolutionTheorem:**Akeypropertyisthatconvolutioninthespatialdomainis equivalent to simple multiplication in the frequency domain, which can be computationally efficient, especially for large kernels.

2DFourierTransform

F(u,v)=∑x∑yf(x,y)e−j2π(ux+vy)F(u,v)=\sum_x\sum_yf(x,y)e^{-j2\pi(ux+vy)}F(u,v)=x∑y∑ f(x,y)e−j2π(ux+vy)

Meaning

- DCComponent(0,0)→Averagebrightness
- Highfrequencies→ Edges
- Lowfrequencies→Smoothareas

SamplingandAliasing

- **SamplingandAliasing**:Samplingistheprocessofconvertingacontinuoussignal(like light entering a camera) intoa discrete digitalimage. Aliasing occurs when a signal is sampled at a rate too low to accurately represent its highest frequency components (below the Nyquist rate). This results in visual distortions like "jaggies" (staircase effectonedges)inimages.Anti-aliasingfilters(typicallylow-passfilters)areused before sampling to remove high frequencies that would otherwise cause this distortion.

FiltersasTemplates

Filtersactas**templates**todetectspecificpatterns.

- **Filters as Templates**: Linear filters, specifically kernels, act as templates. By convolvinganimagewithaspecifickernel,youcanfindpatternsorfeaturesthat match that kernel's design, such as using an edge-detector kernel to find edges

Examples

- Edgedetection kernels
- Linedetectionkernels
- Cornerdetectionkernels

NormalizedCorrelation

Usedfor**patternmatching**.

- **NormalizedCross-Correlation(NCC):**Thistechniqueiswidelyusedforpattern matchingandtemplatematching.Itmeasuresthesimilaritybetweenatemplateand different regions of a larger image.
- **Function:** NCC produces a value between -1(perfect mismatch) and 1(perfect match)andisrobusttolinearchangesinbrightnessandcontrast.Apeakinthe correlationoutputindicatesthelocationwherethepatternismostlikelypresent.

Formula

C=∑(f−f¯)(t−t¯)∑(f−f¯)2∑(t−t¯)2C=\frac{\sum(f-\barf)(t-\bart)}{\sqrt{\sum(f-\barf)^2 \sum(t-\bart)^2}}C=∑(f−f¯)2∑(t−t¯)2∑(f−f¯)(t−t¯)

Range

- −1to+1
- +1→Perfectmatch

ScaleandImagePyramids

- **Image Pyramids**: This is a multi-scale representation of an image, created by repeatedlysmoothing(usuallywithaGaussianfilter)andsubsampling(downscaling) the image. The original, full-size image is at the bottom, and progressively smaller, blurrier versions stack on top.
- **Purpose**:Pyramids areuseful fordetectingobjects thatappearatdifferentsizes or scales within an image, as a fixed-size template can find large objects in lower-resolutionlayersandsmallobjectsinhigher-resolutionlayers.Theyarealsousedin image blending and compression.

ImagePyramid

- Multi-resolutionrepresentation
- Sameimageatdifferentscales

Types

1. **GaussianPyramid**–smoothing+subsampling
2. **LaplacianPyramid**–differencebetweenlevels

Applications

- Objectdetection
- Imagecompression
- Featureextraction

UNIT–II:EDGEDETECTION

NoiseinImages

Additive Stationary Gaussian Noise (AWGN) is a fundamental noise model used in image processingandcommunicationssystems.Itrepresentsrandomdisturbancesthataffectpixel values.

- Additive:Thenoiseissimplyaddedtotheoriginalpixelvalue;it isindependentof the image content.
- Stationary (or White): The noise has a uniform power spectral density across all frequencies,meaningallfrequenciesareequallypresent,analogoustowhitelight.

- Gaussian:Theamplitudeofthenoisefollowsanormal(bell-curve)probability distribution, with a mean of zero and a specific variance (σ2sigma squared

σ2)thatdeterminesthenoisestrength.

- Source:Inpracticalterms,itoftenmodelsthermalnoiseinelectronicsensorsand circuits used for image acquisition.

AdditiveStationaryGaussianNoise

- Noiseis**unwantedrandomvariation**inimageintensity.
- **Additivenoise**:

g(x,y)=f(x,y)+n(x,y)g(x,y)=f(x,y)+n(x,y)g(x,y)=f(x,y)+n(x,y)

- Gaussiannoisefollows**normaldistribution**.
- Stationary→statisticalpropertiesdonotchangeoverspace.

EffectofNoise

- Createsfalseedges
- Reducesaccuracyofedgedetection

WhyFiniteDifferencesRespondtoNoise

Finitedifferenceoperatorsapproximatethederivativeofanimagebycalculatingthe difference between adjacent pixel values.

- Noise Sensitivity: Because noise is oftencharacterizedby high-frequency random variationsbetweenneighboringpixels,thesesimpledifferenceoperationsamplify the noise significantly, creating many false edges.
- SecondDerivatives:Second-orderderivatives(likethebasicLaplacian)areevenmore sensitive to noise than first-order methods (like the basic Sobel operator)
- Finitedifferencesapproximatederivatives:

∂f∂x≈f(x+1)−f(x)\frac{\partialf}{\partialx}\approxf(x+1)-f(x)∂x∂f≈f(x+1)−f(x)

- Noisehas**high-frequencycomponents**
- Derivativesamplifyhighfrequencies→**noiseincreases**

Estimating Derivatives

Tocombatnoisesensitivity,smoothingiscombinedwithdifferentiation.

- Smoothing:AGaussianfilterisappliedfirsttotheimagetosuppresshigh-frequency noise components.
- DerivativeofGaussian(DoG):Insteadofapplying aGaussianfilterandthena derivative operator separately, the operations can be combined due to the associative property of convolution. We can pre-calculate a filter kernel that is the derivativeofaGaussianfunctionandconvolvetheimagewiththissinglekernel.This provides a smoothed derivative estimate, reducing the noise impact while still highlighting intensity changes

First-orderDerivative

- Detects**edges**
- Sensitivetonoise

Second-orderDerivative

- Zero-crossingsindicateedges
- Moresensitivetonoise

DerivativeofGaussianFilters

- Noise Reduction: Smoothing works as a low-pass filter, attenuating the high-frequency components that typically correspond to noise, while preserving importantstructuralinformation(edgesarelowerfrequencythanrandomnoise).
- Gaussian'sOptimality:TheGaussianfilteriswidelyusedbecauseitprovidesan optimal trade-off between spatial localization (accuracy of edge position) and frequencylocalization(effectivenoisesuppression).Itminimizestheproductof spatial and frequency uncertainty, as described by the uncertainty principle.
- Separability:The2DGaussianfilterisseparableintotwo1Dfilters(horizontaland vertical), which makes the computation much more efficient
- Combine**smoothing+differentiation**
- Reducesnoisesensitivity

∂∂x(G∗f)=(∂G∂x)∗f\frac{\partial}{\partialx}(G*f)=(\frac{\partialG}{\partialx})*f∂x∂ (G∗f)=(∂x∂G)∗f

Advantage

- Smoothimage first

- Thencomputederivative

WhySmoothingHelps

- Removeshigh-frequencynoise
- Makesedgedetectionreliable

CommonSmoothingFilter

- Gaussianfilter

ChoosingaSmoothingFilter

- Smallσ→detectsfineedgesbut noisy
- Largeσ→smoothbutmaymiss details
- Trade-offbetween**accuracyandnoisesuppression**

LaplacianforEdgeDetection

TheLaplacianoperatorisasecond-orderderivativeoperator (

∇2f=∂2f∂x2+∂2f∂y2nabla squared f equals partial squared f over partial x squared end-fractionpluspartialsquaredfoverpartialysquaredend-fraction$\nabla 2f = \partial 2f \partial x2 + \partial 2f \partial y2$) that detects edges in all directions.

- Zero Crossings: The Laplacian response is zero in areas of constant intensity and changessign(crosseszero)exactlyatthecenterofanidealedge.Edgesaredetected by finding these "zero crossings."
- Laplacian of Gaussian (LoG): The practical implementation combines Gaussian smoothingwiththeLaplacianoperator(LoGor"Mexicanhat"kernel)tomakeit robust to noise.

Laplacian Operator

∇2f=∂2f∂x2+∂2f∂y2\nabla^2f=\frac{\partial^2f}{\partialx^2}+\frac{\partial^2f}{\partial y^2}∇2f=∂x2∂2f+∂y2∂2f

LaplacianofGaussian(LoG)

- Smoothimage first
- ThenapplyLaplacian
- Edgesdetectedusing**zerocrossings**

Gradient-BasedEdgeDetectors Gradient

Magnitude

|∇f|=(fx)2+(fy)2|\nablaf|=\sqrt{(f_x)^2+ (f_y)^2}|∇f|=(fx)2+(fy)2

GradientDirection

θ=tanθ−1(fyfx)\theta=\tan^{-1}(\frac{f_y}{f_x})θ=tan−1(fxfy)

Examples

- Sobeloperator
- Prewittoperator
- Robertsoperator

OrientationRepresentationsandCorners

Gradient-basedmethodsworkbyfindinglocationswherethefirstderivative(gradient magnitude) of the image intensity is maximum.

- SobelOperator:Acommonexampleuses

3×33cross3

3×3kernelstoapproximatethegradientsinthehorizontal(GxcapGsubxGx)andvertical (Gycap G sub y Gy) directions.

- GradientMagnitude&Orientation:Theedgestrength(magnitude)iscalculatedas

G=Gx2+Gy2capGequalsthesquarerootofcapG subxsquaredpluscapGsubysquared end-root

G=$G2x$+$G2y$

,andthedirection(orientation)is

θ=arctan(Gy/Gx)thetaequalsarctangentopenparencapGsuby/capGsubxcloseparen

θ=arctan(Gy/Gx)

.

- Canny Edge Detector: This robust, multi-stage algorithm uses gradient information, non-maximumsuppression(tothinedges),andhysteresisthresholding(tolinkstrong edges to nearby weak edges).
- Corners:Cornersare pointswhere theimagegradientchangessignificantlyintwoor moredirectionssimultaneously.DetectorsliketheHarriscornerdetectoranalyzethe gradient information to identify these key feature points

- Orientation=directionofedge
- Cornersoccurwhen**intensitychangesinmultipledirections**
- Importantforfeaturedetection

UNIT–III:TEXTURE

WhatisTexture?

Textureisanimportantvisualcuethatdescribesthelocalcharacteristicsofsurfacesinan image. Filter banks are a primary tool for texture analysis.

- FilterBank:Thisisacollectionoffilters designedtoanalyzeanimageatmultiple scalesandorientations.CommonfilterbanksuseGaborfilters,LaplacianofGaussian, or wavelet transforms to capture different frequency and orientation information.
- Process:Convolvingan imagewiththesefiltershelpsreveallocalstructuresbecause a strong filter response occurs where the image pattern matches the filter kernel's design.Thecollectionofresponsesacrosstheentirebankprovidesarichdescription of the local texture

Texturedescribes**surfaceproperties**like:

- Smoothness
- Roughness
- Regularity

Representing TextureUsing FilterBanks

Therawoutputfromalargefilterbankcanberedundant,sostatisticalfeaturesare extracted to create a compact and meaningful texture representation.

- StatisticalFeatures:Foreachfilteredimage(channel),statisticalmeasuresare calculated, such as the mean, variance, skewness, and energy.
- FeatureVectorsandTextons:Thesestatisticsareconcatenatedintoasinglefeature vectorthatrepresentsthetextureofanarea.Insomeapproaches,filterresponses are clustered into prototypical response vectors called "textons," which act as a dictionary for characterizing different texture classes
- Applymultiplefiltersatdifferentorientations&scales
- Eachfiltercapturesspecificpatterns

Example

- Gaborfilters
- OrientedGaussianfilters

StatisticsofFilterOutputs

- Mean
- Variance
- Energy
- Correlation

Usedas**texturefeatures**.

AnalysisandSynthesisUsingOrientedPyramids

- magePyramids:Thesemulti-scalerepresentations(GaussianandLaplacian)arekey to texture analysis and synthesis.
- LaplacianPyramid(LP):Thisstructurecapturesthe"detailimages"orhigh-frequency components at various resolutions by subtracting blurred (low-pass) versions from the original images. The LP is crucial for applications like image compression and blending because it separates details at different scales.
- Oriented Pyramids: These are similar to Laplacian pyramids but incorporate orientation-specificfilters(likeGaborfilters)toanalyzedirectionalstructuresinthe

texture.Theyarehighlyeffectiveforcapturingcomplexpatternsanddirectional information.

- FrequencyDomain: Filterscan bedesignedandanalyzedinthespatialfrequency domainusingtheFourierTransform,whichhelpsinunderstandingtheireffecton different texture frequencies

Analysis

- Decomposeimageintoorientationsandscales

Synthesis

- TextureSynthesis:Thisistheprocessofgeneratingalargedigitalimagefromasmall sampleimagesuchthatthenewimagehasthesamevisualtextureappearance.Itis widely used in computer graphics for tasks like texture mapping in 3D rendering to avoid visible tiling artifacts and seams.
- Homogeneity:Inimageprocessing,homogeneityisameasureofhowsimilarthe elements (pixels ortexels)ofan image region are. Ahomogeneous regionmight consist of the same grey level or a single, uniform texture pattern.
- Synthesis by Sampling Local Models: Many modern texture synthesis algorithms are non-parametricandrelyonsamplingexistingneighborhoods fromasampleimageto create new pixels or patches, often using Markov Random Field models
- Reconstructtexturefromfilterresponses

Laplacian Pyramid

- Multi-scaleimagerepresentation
- Capturestexturedetailsatdifferentresolutions

SpatialFrequencyDomainFilters

- Low-pass→smoothtextures
- High-pass→roughtextures
- Band-pass→periodic textures

OrientedPyramids

- Capturedirectionalinformation
- Usefulfortextureclassification

TextureSynthesis

Goal

- Generatenewtexturesimilartooriginal

Methods

- Samplinglocalmodels
- Statisticalmatching

Homogeneity

- Texture consistencyacrossregions
- Usedinsegmentation

ShapefromTexture

ShapefromTexture(SfT)isacomputervisiontechniquewherea3Dshapeorsurface orientation is recovered from a single 2D image using texture as a primary cue.

- Mechanism:Humanscaneasilyperceivedepthandshapefromtexturechanges(e.g., repeatingpatternsappearingsmalleranddenserfurtheraway).SfTalgorithmsmimic this by analyzing the distortion or perspective changes of texture elements (texels) acrossasurface.
- For Planes: The analysis is often simplified by assuming the surface can be approximatedaslocallyplanar,whichallowsforthederivationofrelationships between texture distortions and surface orientation parameters**.**
- Texturedistortion givesdepth information
- Usedfor3Dsurfaceestimation

UNIT–IV:SEGMENTATIONBYCLUSTERING

WhatisImageSegmentation?

Image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as superpixels). The goal is to simplify and/or change the representationofanimageintosomethingthatismoremeaningfulandeasierto analyze.

- Purpose: The result of image segmentation is a set of segments that collectively covertheentireimage,orasetofcontoursextractedfromtheimage.Eachpixelina region is similar with respect to some characteristic or computed property, such as color, intensity, or texture.
- Applications: Segmentation is a foundational step in most computer vision applications,includingmedicalimaging(tumorboundarydetection),autonomous driving (identifying road from pavement), and object recognition.

Segmentationdividesanimageinto**meaningfulregions**.

HumanVision:GroupingandGestalt

The human visual system automatically groups visual elements into meaningful wholes. ComputervisionalgorithmsoftendrawinspirationfromtheGestaltprinciplesofperception.

- Gestalt Principles: These include proximity (elements close to each other are grouped),similarity(similarelementsaregrouped),continuity(lines/shapesareseen as continuous), closure (incomplete shapes are perceived as complete), and symmetry.
- Relevance:Theseprinciplesinformhowalgorithmsdefine"meaningful"regions during segmentation**.**

Gestaltprinciples:

- Proximity
- Similarity
- Continuity
- Closure

Applications

- Shotboundarydetection
- Backgroundsubtraction
- Objectdetection

ImageSegmentationbyClusteringPixels

Clusteringalgorithmsgrouppixelsbasedontheirproperties(e.g.,intensity,colorvalues,or texture features) in feature space.

- K-MeansClustering:Apopulariterativealgorithmthatpartitionsdatapoints(pixels) into a predefined number (Kcap$K$) of clusters. It is simple and fast but requires knowing the number of clusters beforehand**.**

K-MeansClustering

- Grouppixelsbasedonintensity/color
- Iterativeprocess

SegmentationbyGraph-TheoreticClustering

Graph-basedmethodsmodeltheimageasagraph,wherepixelsarenodesandedgeweights represent the similarity between adjacent pixels. Segmentation then becomes a problem of cutting the graph into meaningful partitions.

- Graph Cuts: Algorithms like normalized cuts find a set of edges to remove (cut) that minimizeacostfunctionrelatedtothedissimilaritybetweensegmentsandmaximize the similarity within segments. These methods often produce high-quality, globally optimal segmentations.

- Pixels=nodes
- Similarity=edgeweights
- Partitiongraphtoformsegments

HoughTransform

TheHoughTransformis afeatureextraction techniqueused in imageanalysis,computer vision,and digitalimage processing. Its primary purpose is tofind imperfect instances of objectswithinacertainclassofshapes(e.g.,lines,circles,ellipses)byavotingprocedure.

- Fitting Lines: It transforms image points (x, y x comma y x,y) into parameter space (e.g., slope mm $m$and intercept bb b, or polar coordinates ρrho ρ and θ the tanθ). Peaksintheaccumulator(parameterspace)indicatethepresenceandparametersof lines in the original image.
- Fitting Curves: The concept extends to more complex shapes like circles or ellipses, thoughtheparameterspacebecomeslargerandcomputationallymoreexpensive.It is a robust method because it canhandle gaps inthe curves orlines and is relatively noise-resistant.

Usedtodetect**linesandcurves**.

LineEquation

ρ=xcosθ+ysinθ\rho=x\cos\theta+y\sin\thetaρ=xcosθ+ysinθ

FittingLinesandCurves

- Robusttonoise
- Usedinlanedetection,shapedetection

UNIT–V:RECOGNITIONBYRELATIONSBETWEENTEMPLATES

ObjectRecognition

This topic addresses how to move beyond simple template matching (like normalized correlationinUnitI)tomorecomplexobjectrecognitionbyconsideringhowpartsofan object relate to each other.

- Part-BasedModels:Objectsareoftenrepresentedascollectionsoffeaturesor "templates"(e.g.,awheel,awindow,adoor)ratherthanasingleholistic template.
- VotingSchemes:MethodsliketheGeneralizedHoughTransformuselocalfeaturesto vote for the probable position and orientation of the entire object's center. This approachisrobusttoocclusion(whenpartsoftheobjectarehidden)becauseevena

fewvisiblepartscanstillcastvotesforthecorrectobjectlocation.Themethod aggregates evidence from different parts to find the best match.

Goal:Identifyobjectsinimages.

VotingonRelationsBetweenTemplates

- Matchpartsinsteadofwholeobject
- Votingaccumulatesevidence

RelationalReasoningUsingProbabilisticModels

To improve the accuracy and robustness of recognition, especially with variations in viewpointandarticulation,weincorporateprobabilisticmodelsandsearchstrategies.

- ProbabilisticModels:ModelslikeBayesiannetworksorMarkovRandomFieldsdefine the probability of finding a part at a specific location given the location of other parts.Thiscapturesthespatialrelationships(e.g.,acar'sleftheadlightistypicallyto the left and slightly above the bumper).
- Search: Recognition often involves searching through a vast parameter space (possiblelocations,scales,orientationsofallparts).Searchstrategies,suchas branch and bound or greedy algorithms, are used to efficiently find the configuration of partsthatbestmatchestheprobabilisticmodeloftheobject.
- Handleuncertainty
- Useprobabilitiestomodelrelationships

SearchandClassifierPruning

Thesearchspacecanbeextremelylarge.Classifiershelpreducethesearcheffortbyquickly discarding unlikely hypotheses.

- HypothesisGeneration&Verification:Initialstepsmightgeneratemanypotential object locations (hypotheses).
- Pruning: Classifiers (e.g., support vector machines, simple neural networks) can be used to rapidly assess the quality of a hypothesis based on local visual features. Hypothesesdeemedhighlyunlikelybytheclassifierarepruned(removed)fromthe search space, allowing the more complex, computationally expensive relational reasoning steps to focus only on promising candidates.

- Reducesearchspace
- Useclassifierstoeliminateunlikelymatches

HiddenMarkovModels(HMM)

HiddenMarkovModelsarepowerfulstatisticalmodelsusedtomodelsystemsthatcanbein certain "hidden"states (whicharenotdirectly observable) and produce observableoutputs based on those states.

- Sequential Data: HMMs are particularly effective for sequential or temporal data, suchasspeech,handwriting,orgesturesinsignlanguage.Thesequenceofobserved images(features)correspondstoasequenceof underlyinghiddenstates(theaction being performed).
- Training & Recognition: During training, the model learns the probabilities of transitioning between states and the probabilities of observing specific features in eachstate.Duringrecognition,itusesalgorithmsliketheViterbialgorithmtofindthe most likely sequence of hidden states (e.g., the specific sign being performed) given an input sequence of images.
- Statisticalmodelforsequentialdata

Components

- States
- Observations
- Transitionprobabilities
- Emissionprobabilities

ApplicationsofHMM

- **SignLanguageUnderstanding**:HMMsareaclassicapproachforrecognizingdynamic gestures and signs. The observed visual features (hand position, joint angles, movementtrajectoryovertime)aremodeledasoutputsfromhiddenstates representing different phonemes or signs.
- **Finding People with HMM**: HMMs can be adapted for spatial tasks too, such as modeling the typical visual sequence found when scanning vertically for a person (e.g.,headfeatures->torsofeatures->legfeatures).Themodelcanslideawindow across an image and use the HMM to evaluate if the vertical sequence of features matches a typical human profile structure**.**
- Speechrecognition

- Gesturerecognition
- Signlanguageunderstanding
- Findingpeopleinvideos