18ECE340T - MACHINE PERCEPTION WITH COGNITION

Unit - IV

Template Matching:

The process of template matching is done by comparing each of the pixel values of the source image one at a time to the template image. The output would be an array of similarity values when compared to the template image.

(any relevant EXAMPLE can be taken)

Template matching can also be used as a pipeline in conducting object detection for machine learning models and deep learning models.

Template matching is a high-level computer vision approach that detects image portions that match a predetermined template. Advanced template matching algorithms detect template occurrences regardless of orientation or local brightness.

In medical image analysis, invariant characteristics or innovative applications are commonly used as object identification domains such as vehicle tracking, robotics, and manufacturing.

Template Matching approaches are versatile and simple to apply, making them one of the most used ways of object localization. Their utility is restricted mostly by computer capacity, as identifying large and complicated templates can be time-consuming.

- It is a way of searching for and locating a template within a larger image.
- The goal is to discover identical portions of an image that match a template we provide based on a threshold.
- The threshold determines how well we want to recognize the template in the original image.
- For example, if we are using face recognition and want to detect a person's eyes, we can use a random image of an eye as the template and search for the source (the face of a person).
- In this scenario, because "eyes" vary so much from person to person, even if we set the threshold to 50% (0.5), the eye will be recognized.

Working Function

- Simply slide the template picture over the input image (as in 2D convolution)
- The template picture and the piece of input image beneath it are compared.
- The acquired result is compared to the threshold.
- If the result exceeds the threshold, the section is marked as detected.

- It is a method for locating a reference image (or a template image) within a source image. In its most basic form, the algorithm compares the template for each source image region, one pixel at a time. This is referred to as cross-correlation. The result of this procedure is another image with a pixel value that corresponds to how similar the template image was to the source image when it was inserted at that pixel location.

(Refer PPT)

Hidden Markov Model:

HMM is a statistical **Markov model** in which the system being modeled is assumed to be a **Markov process** with unobserved (**hidden**) states.

Markov Model: *Series of (hidden) states z={z_1,z_2…………} drawn from state alphabet S ={s_1,s_2,…….s_|S|} where z_i belongs to S.*

Hidden Markov Model: *Series of observed output x = {x_1,x_2,………} drawn from an output alphabet V= {v1, v2, . . . , v_|v|} where x_i belongs to V*

Assumptions of HMM

HMM too is built upon several assumptions and the following is vital.

- **Output independence assumption**: Output observation is conditionally independent of all other hidden states and all other observations when *given the current hidden state.*

$$P(x_t=v_i / z_t=s_j) = P(x_t=v_i / x_1, x_2, .., x_T, z_1, z_2, .., z_T) = B_{ji}$$

Eq.5.

- **Emission Probability Matrix**: Probability of hidden state generating output v_i given that state at the corresponding time was s_j.

Hidden Markov Model as a finite state machine

Consider the example given below in Figure. which elaborates how a person feels on different climates.

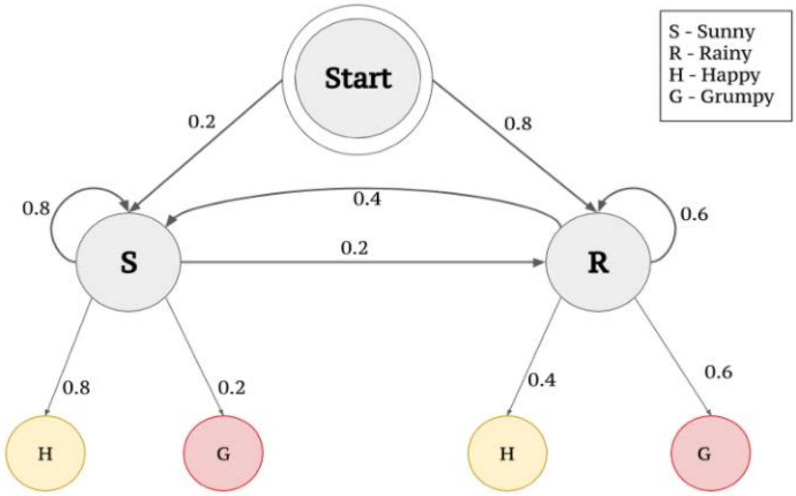


Figure: Markov Model as Finite State Machine

Set of states (S) = {Happy, Grumpy}

Set of hidden states (Q) = {Sunny , Rainy}

State series over time = z∈ S_T

Observed States for four day = {z1=Happy, z2= Grumpy, z3=Grumpy, z4=Happy}

The feeling that you understand from a person emoting is called the observations since you observe them.

The weather that influences the feeling of a person is called the hidden state since you can't observe it.

Emission probabilities

In the above example, feelings (Happy or Grumpy) can be only observed. A person can observe that a person has an 80% chance to be Happy given that the climate at the particular point of observation(or rather day in this case) is Sunny. Similarly the 60% chance of a person being Grumpy given that the climate is Rainy. Here mentioned 80% and 60% are Emission probabilities since they deal with observations.

Transition probabilities

When we consider the climates (hidden states) that influence the observations there are correlations between consecutive days being Sunny or alternate days being Rainy. There is 80% for the Sunny climate to be in successive days whereas 60% chance for consecutive days being Rainy. The probabilities that explain the transition to/from hidden states are Transition probabilities.

Three important questions in HMM are

What is the probability of an observed sequence?

What is the most likely series of states to generate an observed sequence?

How can we learn the values for the HMMs parameters A and B given some data?

1. Probability of Observed Sequence

We have to add up the likelihood of the data x given every possible series of hidden states. This will lead to a complexity of O(|S|)^T. Hence two alternate procedures were introduced to find the probability of an observed sequence.

Forward Procedure

Calculate the total probability of all the observations (from t_1) up to time t.

$\alpha\_i\ (t) = P(x\_1\ ,\ x\_2\ ,\ \ldots\ ,\ x\_t,\ z\_t = s\_i;\ A,\ B)$

Backward Procedure

Similarly calculate total probability of all the observations from final time (T) to t.

β_i (t) = P(x_T , x_T-1 , …, x_t+1 , z_t= s_i ; A, B)

Background Subtraction:

Background subtraction is a way of eliminating the background from image. To achieve this, the moving foreground is extracted from the static background. This technique is used for detecting dynamically moving objects from static cameras. Background subtraction technique is important for object tracking.

The base in this approach is that of detecting moving objects from the difference between the current frame and reference frame, which is often called ‘Background Image’ or ‘Background Model’. This background subtraction is typically done by detecting the foreground objects in a video frame and foreground detection is the main task of this whole approach.

Many applications do not need to know the whole contents of the sequence, moreover, further analysis is focused on some part of the sequence because interest lies in the particular objects of images in its foreground. After completing all the pre-processing steps such as deionizing, morphological processing, object localisation is carried out and there this foreground detection is used.

All the present detection techniques are based on modelling the background of the image i.e. set the background and detect the changes that occur. Defining the proper background can be very difficult when it contains shapes, shadows and moving objects.

While defining the background, it is assumed by all the techniques that the stationary objects could vary in colour and intensity over time.

Background Subtraction has several use cases in everyday life, it is being used for object segmentation, security enhancement, pedestrian tracking, counting the number of visitors, number of vehicles in traffic etc. It is able to learn and identify the foreground mask.

Histograms Introduction

In digital image processing, the histogram is used for graphical representation of a digital image. A graph is a plot by the number of pixels for each tonal value. Nowadays, image histogram is present in digital cameras. Photographers use them to see the distribution of tones captured.

In a graph, the horizontal axis of the graph is used to represent tonal variations whereas the vertical axis is used to represent the number of pixels in that particular pixel. Black and dark areas are represented in the left side of the horizontal axis, medium grey color is represented in the middle, and the vertical axis represents the size of the area.

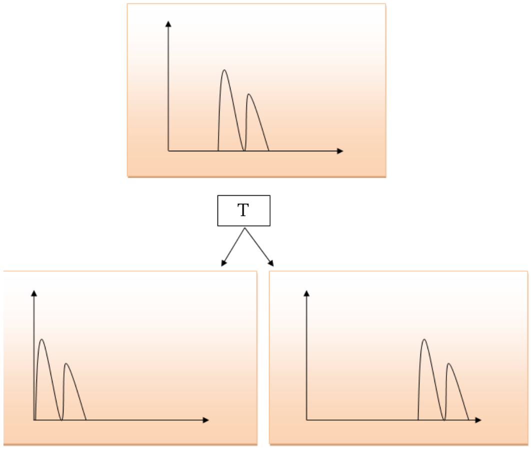
Applications of Histograms

- In digital image processing, histograms are used for simple calculations in software.
- It is used to analyze an image. Properties of an image can be predicted by the detailed study of the histogram.
- The brightness of the image can be adjusted by having the details of its histogram.
- The contrast of the image can be adjusted according to the need by having details of the x-axis of a histogram.
- It is used for image equalization. Gray level intensities are expanded along the x-axis to produce a high contrast image.
- Histograms are used in thresholding as it improves the appearance of the image.
- If we have input and output histogram of an image, we can determine which type of transformation is applied in the algorithm.

Histogram Processing Techniques

Histogram Sliding

In Histogram sliding, the complete histogram is shifted towards rightwards or leftwards. When a histogram is shifted towards the right or left, clear changes are seen in the brightness of the image. The brightness of the image is defined by the intensity of light which is emitted by a particular light source.

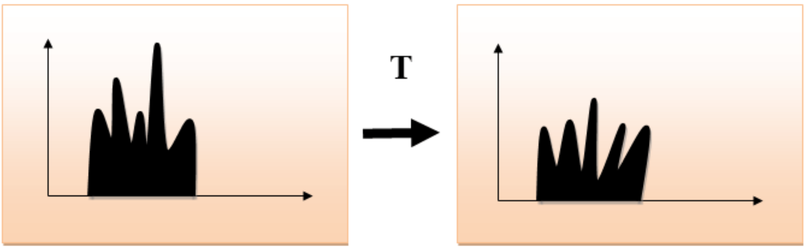


Histogram Stretching

In histogram stretching, contrast of an image is increased. The contrast of an image is defined between the maximum and minimum value of pixel intensity.

If we want to increase the contrast of an image, histogram of that image will be fully stretched and covered the dynamic range of the histogram.

From histogram of an image, we can check that the image has low or high contrast.

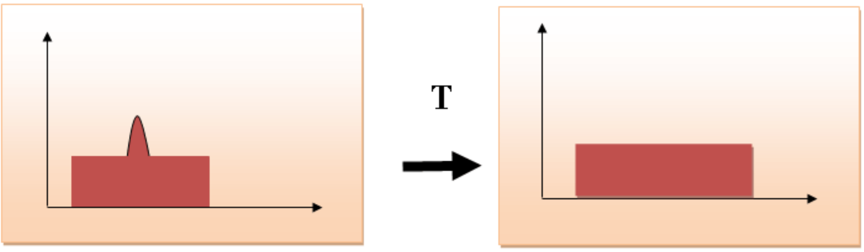


Histogram Equalization

Histogram equalization is used for equalizing all the pixel values of an image. Transformation is done in such a way that uniform flattened histogram is produced.

Histogram equalization increases the dynamic range of pixel values and makes an equal count of pixels at each level which produces a flat histogram with high contrast image.

While stretching histogram, the shape of histogram remains the same whereas in Histogram equalization, the shape of histogram changes and it generates only one image.



Refer for detailed information

Hough Transform

Line detection algorithm: https://theailearner.com/2020/11/20/hough-line-transform/

Circle detection algorithm: https://theailearner.com/tag/hough-circle-transform-algorithm/

Hidden Markov's Model (HMM): https://towardsdatascience.com/markov-and-hidden-markov-model-3eec42298d75

MCQs link for reference (not limited to):

Template Matching: https://mcqmate.com/discussion/6286/which-of-the-following-involves-correlation (at the bottom check for related questions)

https://mcqmate.com/discussion/6268/what-is-the-basis-for-numerous-spatial-domain-processing-techniques

HMM: https://engineeringinterviewquestions.com/mcqs-on-hidden-markov-model-and-answers/

https://www.sanfoundry.com/artificial-intelligence-mcqs-hidden-markov-model/

Unit - V

Human Gait Analysis

Gait is the action of walking (locomotion). It is a complex, whole-body movement, that requires the coordinated action of many joints and muscles of our musculoskeletal system. It mostly includes the movements of the lower limbs, upper limbs, pelvis and spine.
Gait also depends on the proper functioning of other body systems such as nervous, cardiovascular and respiratory system.

Phases of gait

The two main phases of gait include:

- The stance phase
- The swing phase

Each of these phases has subphases that are going to be described in detail in the following text.

Stance phase

The stance phase is the period of the gait cycle when the foot is on the ground and bearing body weight. More specifically, it can be described as the period between the moment that the heel of the foot touches the ground (heel strike) until the moment that the toe-off occurs. The stance phase consists of five subphases;

1. The heel strike (initial response, contact response, or weight acceptance). In this subphase, the heel of the foot makes initial contact with the ground. It requires the body's weight to be accepted by the leg making contact with the ground.
2. The foot flat (loading response) is the second subphase when the foot rolls forward until the entire plantar surface is in contact with the ground.
3. The midstance starts when the weight of the body is propelled forward, directly over the lower extremity, so that the greater trochanter of the femur is directly above the middle of the foot. At this stage, our entire body weight is being balanced over one leg.
4. The next substage is heel-off and includes lifting the heel off the ground. This is when we start to shift the body weight onto the contralateral leg.
5. The toe-off is the final stage of the stance phase and includes pushing the toes into the ground while the ankle plantarflexes, creating forward propulsion.

Swing phase

The swing phase is the second phase of gait when the foot is free to move forward. It is described as the period between toe-off and heel strike. There are three subphases of the swing phase:

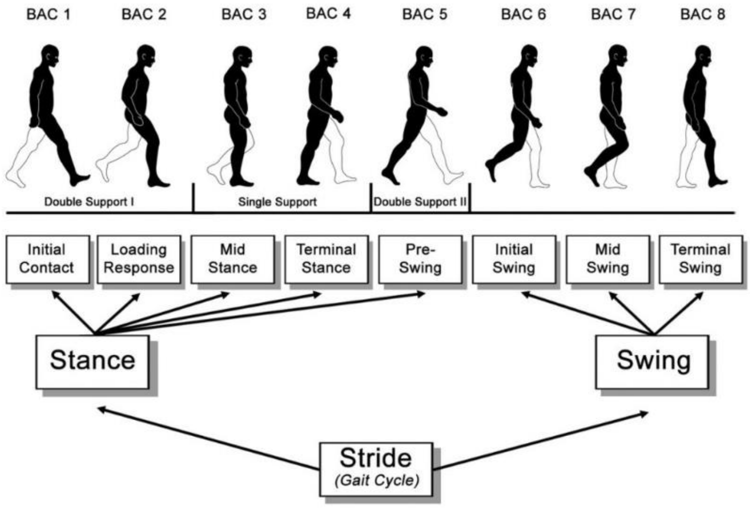
1. The early swing (acceleration phase) is the first sub-phase during which the foot is lifted from the ground. The ankle dorsiflexes and the knee flexes so that the foot and toes can be moved from the ground. The hip flexes to bring the leg forward, moving it directly under the body.
2. The mid-swing phase is the second phase when the non-weight-bearing leg passes directly beneath the body and past the weight-bearing leg. At the same time, the trunk is moved forward so that the weight of the body is directly over the weight-bearing leg.
3. The late swing (deceleration phase) is the last subphase. The foot is moved to a position in front of the body, the knee extends and momentum decelerates. The lower limb is now ready for heel strike and prepares to accept the transfer of body weight, for the start of the next stance phase.

As the legs move, so does the rest of the body. The pelvis rotates forward along with the swing leg, while at the same time the thorax and spine rotate in the opposite direction to maintain balance. The contralateral arm assists balance and propulsion, by swinging forward at the same time as the opposite leg.

Gait is a multi-joint, multi-organ activity. Muscle contraction and joint mobility are considered to be key elements in the functioning of gait.

Muscles may contract concentrically (muscles shorten, creating movement and acceleration), eccentrically (controlled muscle lengthening, creating deceleration) or isometrically (muscle contracts without shortening or lengthening, creating stability).

Figure 1. Breakdown of the gait cycle into phases based on the work of Perry and Burnfield (2010)



(Refer PPT for diagrams and Cycles)

MCQ: https://www.proprofs.com/quiz-school/story.php?title=msak-gait-analysis

Combining Views from Multiple Cameras

Using multiple cameras to shoot the same subject were first done on television in the 1920s. Cinerama is a widescreen process that use three film cameras and then three film projectors.

Combining pictures into one picture is called Compositing.

If by precision (i.e.) its resolution, then the process is called Multi-exposure Image Noise Reduction.

If by precision (i.e.) its dynamic range/luminance, then the process is called High Dynamic Range Imaging.

If by precision (i.e.) freezing the subject's motion, then the process is called Timeslicing.

Taking shots of the same subject with different camera settings, by one or more cameras, is called Bracketing.

(Refer PPT to include diagram and descriptions)

Multiple Camera Object Tracking

Object Tracking : It is to obtain an accurate estimate of the position (x,y) of the object tracked. Tracking algorithms can be classified into:

- Single object Single Camera
- Multiple object Single Camera
- Multiple objects Multiple Cameras
- Single object Multiple Cameras

Single Object Single Camera

- Accurate camera calibration and scene model
- Suffers from Occlusions
- Not robust and object dependant

Single Object Multiple Camera

- Accurate point correspondence between scenes
- Occlusions can be minimized or even avoided
- Redundant information for better estimation
- Multiple camera Communication problem

Static Point Correspondence

- The output of the tracking stage is a simple scene model is used to get real estimate of coordinates
- Both Affine and Perspective models were used for the scene modeling and static corresponding points were used for parameter estimation
- Least mean squares was used to improve parameter estimation

Block-Based Motion Estimation

- Typically, in object tracking precise sub-pixel optical flow estimation is not needed.
- Furthermore, motion can be on the order of several pixels, thereby precluding use of gradient methods.
- We started with a simple sum of squared differences error criterion coupled with full search in a limited region around the tracking window.

Adaptive Window Sizing

- Although simple block-based motion estimation may work reasonably well when motion is purely translational, it can lose the object if its relative size changes.
- If the objects camera field of view shrinks, the SSD error is strongly influenced by the background.
- If the objects camera field of view grows, the window fails to make use of entire object information and can slip away.

Four Corner Method

- This technique divides the rectangular object window into 4 basic regions - each one quadrant.
- Motion vectors are calculated for each subregion and each controls one of four corners.
- Translational motion is captured by all four moving equally, while window size is modulated when motion is differential.
- Resultant tracking window can be non-rectangular, i.e., any quadrilateral approximated by four rectangles with a shared center corner.

Correlative Method

- Four corner method is strongly subject to error accumulation which can result in drift of one or more of the tracking window quadrants.
- Once drift occurs, sizing of window is highly inaccurate.

- Need a method that has some corrective feedback so window can converge to correct size even after some errors.
- Correlation of current object features to some template view is one solution.
- Basic form of technique involves storing initial view of object as a reference image.
- Block matching is performed through a combined interframe and correlative MSE where sc(x0,y0,0) is the resized stored template image.
- Furthermore, minimum correlative MSE is used to direct resizing of current window.

Occlusion Detection

- In order for multi-camera feature tracking to work, each camera must possess an ability to assess the validity of its tracking (e.g. to detect occlusion).
- Comparing the minimum error at each point to some absolute threshold is problematic since error can grow even when tracking is still valid.
- Threshold must be adaptive to current conditions.
- One solution is to use a threshold of k (constant gt 1) times the moving average of the MSE.
- Thus, only precipitous changes in error trigger indication of possibly fallacious tracking.

Redetection Procedure (1 Camera)

- Redetection is difficult at most general level Object recognition.
- Proximity and size constancy constraints can be imposed to simplify redetection.

finally, Multiple cameras can do more than just 3D imaging Camera calibration only works if you have an accurate scene and camera model

- Tracking is sensitive to the camera characteristics (noise, blur, frame rate,..)
- Tracking accuracy can be improved using multiple cameras

Face Detection and Face Recognition

Face Detection: The face detection is generally considered as finding the faces (location and size) in an image and probably extract them to be used by the face detection algorithm.

Face Recognition: The face recognition algorithm is used in finding features that are uniquely described in the image. The facial image is already extracted, cropped, resized, and usually converted in the grayscale.

Face recognition is a solution designed to recognize a human face without any physical contact required. The system runs through algorithms that match the facial node of a person to the images saved in a database. The system first recognizes the unique features such as eyes, ears, nose,…etc, and compares those features to the images stored in the database.

When the features of the person match with the image then the system shows that the person is matched with that particular image. Face recognition systems are used to find the identification of the criminals, to track the attendances in schools, colleges, etc.

There are two types of comparisons are there in face recognition like verification & identification. The four stages to identify the person's face are capture, extraction, comparison & match, or no match. The components of this system are the enrolment module, identification module & database.

Distance between the eyes, Width of the nose and Cheekbones, jawline, chin are the nodal points measured by the software for face recognition. The implementation of face recognition technology include: Data acquisition, Input processing, decision making and Face image classification.

Face recognition algorithm:

- Principal components analysis
- Linear discriminant analysis
- Elastic bunch graph matching

The accuracy of the face recognition system is approximately 92% whereas its accuracy will decrease based on the lighting condition, make up & face position. These systems are used for criminal identification, tracking attendance, surveillance, etc. Other difficulties in face recognition include head pose, aging problem and Occlusion,

MCQ: https://www.watelectronics.com/mcq/face-recognition/