

Biometrics
Prof. Phalguni Gupta
Department of Computer Science and Engineering
Indian Institute of Technology, Kanpur

Lecture No. # 12

(Refer Slide Time: 00:28)

Matching

- Matcher is a system that takes two samples of biometric data and returns a score that indicates their similarity or dissimilarity.
- Similarity or dissimilarity measures are highly dependent on the
 - *data acquisition device,*
 - *precision of representation of biometric samples,*
 - *degree of uniqueness* of the samples over time, etc.

So, we will be discussing on basic system errors and see what happens in the case of biometrics as we have mentioned there are few components, one component is matcher. Of course, there exists the other components it is image acquisition, pre processing, feature extraction and other issues are there, but here this error is dependent on many factors. But matcher is the primary component which will help you to take the decision.

So, given the two biometrics characteristics of same person or may be two different persons, the matcher will tell you that whether it is matched or not. That means what that matcher will give you some information about the similarity or dissimilarity.

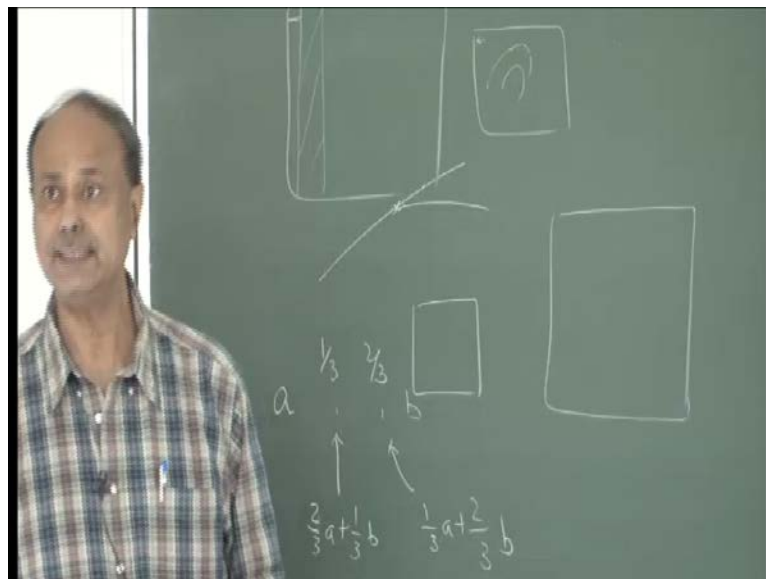
Now, as you know that we will be using the features from the biometrics characteristics, so this, while you do the sum matching operations it will give you some scope, that scope based on that scale scope you have to tell it is matched or not. What it means that there is a need of defining some threshold, that if it is above some threshold in the case of similarity match or below some threshold in the case of dissimilarity match, you will

tell that the two biometrics characteristics are from the same persons that means you will be telling that these two are matched.

So, the word you will find the similarity and the dissimilarity say for example, I have the fingerprint; in the fingerprint we have used the terminology minutiae points and minutiae points is nothing but that you know this ridges are there, their end points or bifurcation points.

Now, given 2 minutiae points what is your given 2 biometrics characteristics of fingerprints, if you I tell you that you matched or not what you do you'll be abstractive the features. Now this feature is the region points or bifurcation points. Now, this end point is represented by x y coordinate and also the angle - angle with respect to the this line, then you will find out that what angle it is creating and also in the case of bifurcation that you can think that, this is one point and then bifurcation looks like this.

(Refer Slide Time: 03:14)



So, this a they are basically you will be getting this points and also that at this point what is the angle. So, there are four parameters one is the type of the minutiae point, one is the x coordinate, another one is y coordinate, another one is the angle. Now, based on this given the two main fingerprints you will be finding how many of the minutiae points are matched and the number is more then you tell it they are similar or they are matched.

So, out of say 80 minutiae points suppose you got the 60 minutiae points matched between the two biometrics characteristics, then you will be telling obviously you will be telling that they are matched, but if it is 50 are you going to tell matched that is not

known to you, so based on your training data you will be taking the decision and this is an example of similarity match.

Now, this similarity or dissimilarity measures are dependent on many parameters; one is that data acquisition system that what type of system you are using, how robust the system it is; all those things you have to take into account. For example, that you I have a fingerprint scanner which works under control environment, will it work if the temperature is more than 50 degree centigrade or will it work if it the temperature is minus 5 degree and so on.

Or that is it that, every system has certain life period so I have a system fingerprint scanner which I have purchased 10 years back, will it work for next 15 years? So, the device plays an important role and then the scanner quality is getting change, say today in the market I have ten megapixel camera, but few years back I had only two megapixel based camera so data acquisition system plays an important role.

Then the precision of the representations, see I have the fingerprint this is one inch by one inch shape and there are 100 points in it. So, obviously this points will be very closely as so closely placed, the coordinate point are very nearby area. Now I have another fingerprint they are also having 100 points, now suppose I consider only the integer value x and y are the integer coordinates then you will find all of them are matching or most of them are matching.

So, you have to think about the no after decimal points also you have to consider, but that representation you have to check what should be the precision, how many digits after decimal point, you have to consider that depends on the biometrics characteristics you are considering. So, precision of representation of biometrics sample is important parameters.

How many minutiae points you are looking for, this is also an important parameter. Suppose, I have a biometrics characteristics we are defining and it gives you only 4 or 5 minutiae points, is it sufficient to draw a conclusion whether two or them are matched or not. So, this is the second parameter. Third one is that degree of uniqueness, whatever you are taking features are the unique, how long you it this property should be maintained that has to be started carefully, so degree of uniqueness is also important parameters. So, the similarity and dissimilarity are mainly dependent on this parameters.

So, you have to be little careful while designing your matching algorithms. See I have a face image, which has been obtained from a camera using a camera of having the resolution of two megapixel and I have another camera, now which has the resolution of say 12 megapixels. Now if I take 12 megapixel based used the 12 megapixel based camera and the size will be very large.

Now, what are you going to do obviously you have to make in the same format same size, suppose I may get 100 crossed 250 pixels image size then there is a possibility you will be losing certain features from the frame because you know resizing how are you going to resize that. One way is that suppose I have one image and another image like this, so one way you will be doing that you will be make because you want to compare this two image. So, this size has to be same otherwise, it will be difficult for you to compare one way that you increase this one or zoom this image to a larger area.

Now, as you have seen in your mid term exam paper that zooming is not that easy, even though its looks very simple.

Have you done that?

(O)

How is it? How did you do it?

(O)

Three times, so what it gives you, that have you done you write a simple program and I did three times you will find there will be of a square effect is coming. Square effect means, this was the image and you have zoomed like this so you will find that this pixel you have repeated, same effect will be there. That, this pixel you have repeated that you will find that there is a shade is coming, some shade will be coming square, square, square shade will come, these are good practice.

Simple way is that, you take the two pixel make it average of it is better, but what I told know if you average it, then you get the two types but what you wanted do not know two times you have want three times, that is the only problem I get.

So, basically the problem you should have written suppose this pixel value is a, this pixel value is b and you want to insert this two points, that is the problem I told you and you know this is one third and this is two third. Distance from here it is one third and this is for two third. So, the value of this should have been two third of a, plus one third of b

and this should have been one third of a, plus two third of b that is the better estimate, because I am giving more weightage here, because in this point it is nearer to a, this point is nearer to b, so I am giving the more weightage on b, that is a simple linear interpolation technique.

So, the zooming even it look very simple the repetitor you should be very careful while answering the by seeing the question paper, you understood the paper is lengthy and once you are lengthy there are some questions which are very simple in nature, this is very simple in nature.

Then another one is that, even though I changed the orientation that it is a orientation free something I gave invariant, del square you have to see that x or that corresponding change and also you take the second derivative. They are same that is the thing, you have to show that is also very half page or one page, half paged item you could have solved straightway there three four problems straightway, you earn the marks and there are some problem where you have to do little arithmetic, simple arithmetic they are lengthy anyway.

Now, the I have one image of two pixel size two megapixel, another image based on the twelve megapixels. Now if you have to compare it, you have to one you have to what you have to one has to meet of other size. The zooming is one way, but by putting the zooming you may introduce some additional features say for example, in the case of fingerprint I, if I zoom it then there is a possibility, there some bifurcation or some additional n points will come.

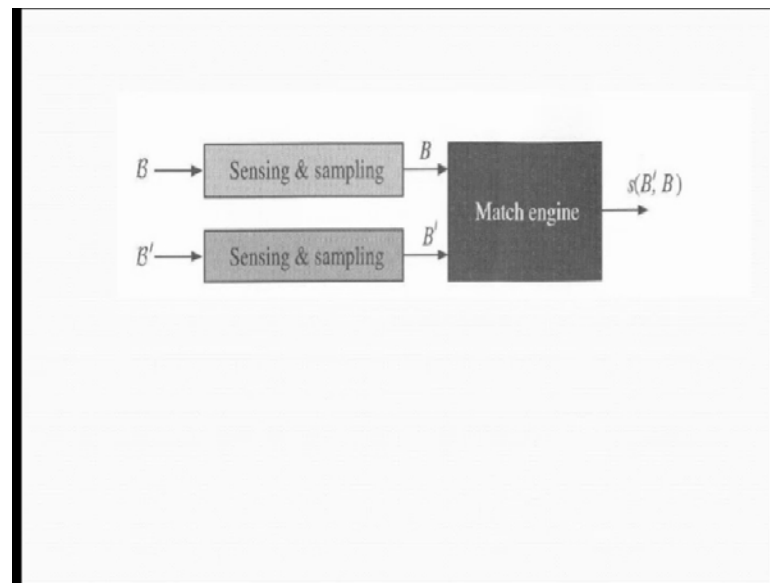
See what happens once you give the fingerprint, once you give the fingerprints the ridges are like this. Once you give the fingerprints, so this is a possibility because of some noise there are some cut is there. Now, if you enlarge it then this cut also will be enlarged or so you need to have, because obviously if I see it minutely this cannot be a cut, this is because of noise only you got that blank area. So, you have to feel that and make it a line, but by introducing the drawing and line I could have done this one I will increase because I will assume that this is line, if the number of pixel where it becoming 0 or it is not line. So, I will add no row, it is a by because of error so I will add that one.

But by introducing that line again, there is a possibility of making an error because in practice actually there is a line, because one line one pixel that may be there because you

cannot ask that concern subject, yes you come and show your fingerprint based on that I will link or I will not link that is not possible.

So, if you zoom this image or then 100 percent sure that line break will be shown more prominently, but if you shrink the image, shrinking then what that line break may go, so that is lot of issues are there. So, this enlarging this one is not that easy or reduce suppose, if we reduce it that may be two features, correct features you may lose right. So, while designing the matchers you must ensure that device quality or device performance or device characteristics are taken into account correctly.

(Refer Slide Time: 15:13)



So, this is the simple thing that you have the biometric subject, another subject has come, may be same subject they have given the data through sensors and the sensors extracted the features and then through matching engine, he gives you some scope

(Refer Slide Time: 15:34)

Two Kinds of Errors

Given two biometric samples, we can have two possible hypotheses:

Null Hypothesis $H_0 \Rightarrow$ Two samples match
Alternate Hypothesis $H_1 \Rightarrow$ Two samples do not match

Definitions of hypotheses depend on biometric applications
 \Rightarrow Different applications can have different definitions of errors

But matching engine decides either H_0 is true or H_1 is true

So, if you know this two biometric samples in front of you, then obviously you will be able to design or you have to straight the two hypotheses. These are the two possible hypotheses, one is null hypotheses and another one alternative hypotheses. In the case of null hypotheses, you will be telling that two samples are matched and in case of alternative hypotheses you will be telling that two samples have not matched.

Now, this definition of hypotheses is again dependent on the applications for what purpose, you when are you going to tell that two are matched. In my classroom out of suppose, I introduce a fingerprint scanner to take your attendance or as you know that in front of you have one fingerprint scanner you are using, is there anybody who face the problem, that not enrolled or not accepted.

Did you try are you using giving your attendance?

(O)

No.

You are doing this attendance in front that kiosk is there card and then you give the fingerprint?

(O)

No, just enter the card only.

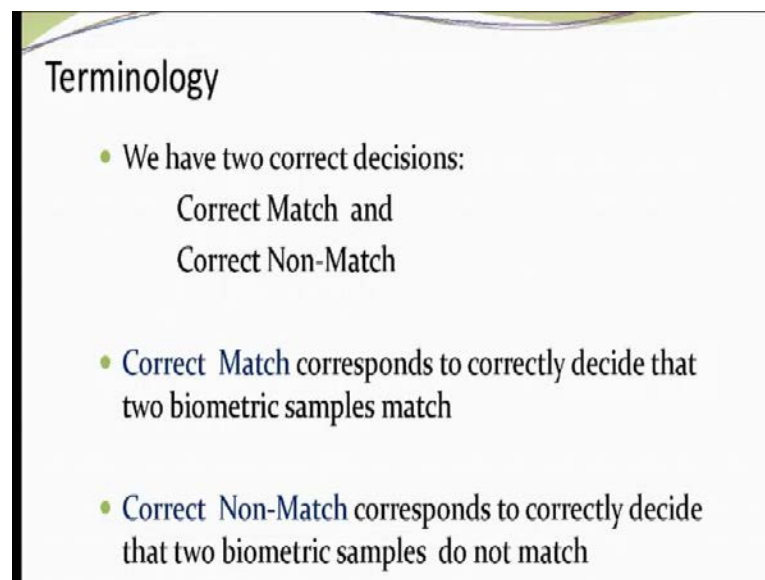
See, what happens that in my attendance system classroom attendance system suppose, I get the 80 minutiae points and out of 80 minutiae points even if I find that 5 of them are

matched, I will tell that attended. Because I, it is not the gain thing that somebody else will know I am sure, somebody else will not come to my class for why did he does not gain something from me or either you know, there people they come to the class to earn 5 percent marks as you are attending.

There are people, very few people will be attending any function even if he is it is not of his interest. There may be something I will attend a seminar, if it is of my field this is simple thing. So, if you tell no you have to attend may be you will be attending, but you will be sleeping nothing will come out of it. So, this in my hypotheses - null hypotheses dependent on the application what purpose I am designing. Suppose, I am thinking that no this attendance is very important, then obviously I will put that out of 80 minutiae points, 70 minutiae points must be matched. So, this will be little tighter side.

Now, this applications again is dependent on the different types of error as it I told you that a 5 out of 80 if it is matched, I will tell yes he attended. Another case is that, no 70 out of 80 if it matched then only I will tell he attended. So, this error range will be different but by matching a engine must return one of them either you have to tell do not match or you have to tell that match. So, one of the hypotheses true either you have to tell H_1 is true or H_0 is true, you cannot tell we do not know which one is true.

(Refer Slide Time: 19:07)



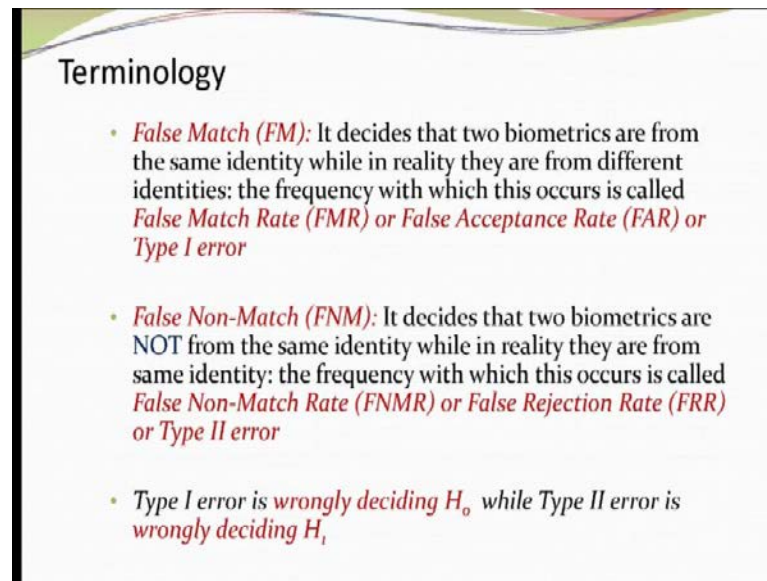
Terminology

- We have two correct decisions:
Correct Match and
Correct Non-Match
- **Correct Match** corresponds to correctly decide that two biometric samples match
- **Correct Non-Match** corresponds to correctly decide that two biometric samples do not match

So, you get the two term obviously one is the correct match, another one term will be a correct not match. Correct match means, that you have got the two biometric samples and they are correctly telling you that, yes these two samples are from the same persons

and correct not match means that there you got the two samples from the two different identities or two different subjects and your system must be able to tell correctly. They are two from the two subjects and the ideal situation is that under every circumstance for genuine case, it should give you always the correct match and for all the imposter case, you should give always correct not match that is the ideal scenario.

(Refer Slide Time: 20:11)



Terminology

- *False Match (FM)*: It decides that two biometrics are from the same identity while in reality they are from different identities: the frequency with which this occurs is called *False Match Rate (FMR) or False Acceptance Rate (FAR) or Type I error*
- *False Non-Match (FNM)*: It decides that two biometrics are NOT from the same identity while in reality they are from same identity: the frequency with which this occurs is called *False Non-Match Rate (FNMR) or False Rejection Rate (FRR) or Type II error*
- *Type I error is wrongly deciding H_0 while Type II error is wrongly deciding H_1*

But this is not the case and as a result there are two terms are coming, type one and type two error. This is true for any pattern recognition problem and here also it is biometrics also based on the pattern recognition. So, these two errors will be there, one term is false match given the two biometrics characteristics you will be telling yes they are matched, but in reality that should not be happen - should not have happened they are two or this characteristics have been obtain from the two different subjects.

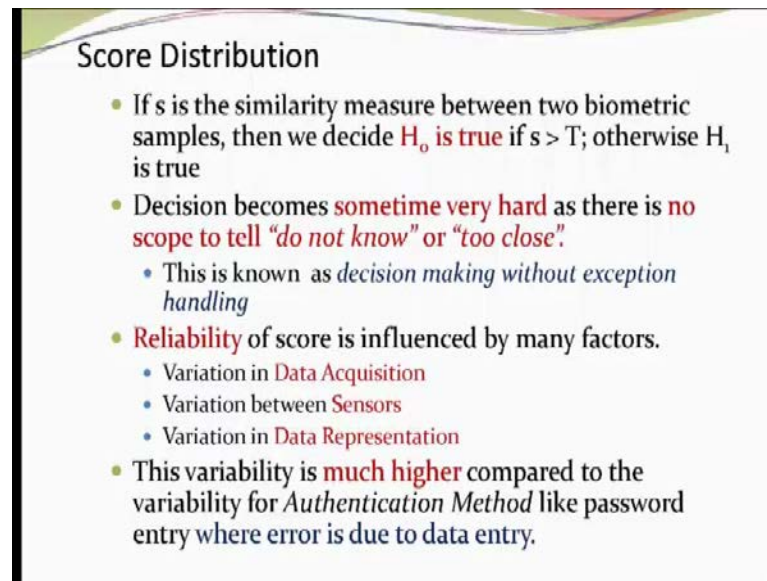
Now, the frequency of that type of error is known as false match rate or false acceptance rate or type one error. In our case, most of the places you will find that we are using FAR false acceptance rate. So, false acceptance is the ratio of number of times it fail or number of times it falsely matched, falsely accepted divided by total number of times you have tested.

Suppose you have tested n times, out of n times k times falsely matched then that will become the ratio, k by n in to 100 that will give you the percentage and false non match is the it decides the two biometrics are not from the same subjects, but in reality they are from the same subjects. You are telling, no they are two different subject but in reality

they are from the same subject and the total frequency of occurrence such type of error is known as false not match rate or false rejection rate and type two error.

So type one, a type of error gives you the indication that your H_0 is true wrongly. Type one error is responsible for wrong acceptance of H_0 and similarly type two error is responsible for the wrong acceptance of H_1 or rejection of H_0 .

(Refer Slide Time: 22:44)



Score Distribution

- If s is the similarity measure between two biometric samples, then we decide H_0 is true if $s > T$; otherwise H_1 is true
- Decision becomes **sometime very hard** as there is **no scope to tell "do not know" or "too close"**.
 - This is known as *decision making without exception handling*
- **Reliability** of score is influenced by many factors.
 - Variation in **Data Acquisition**
 - Variation between **Sensors**
 - Variation in **Data Representation**
- This variability is **much higher** compared to the variability for *Authentication Method* like password entry where error is due to data entry.

So, s is the score matching score between the two biometrics template this word also you will be finding that very frequently, we will be using the template, template is nothing but the feature vector. That means, what their feature vector is nothing but that vector containing the features obtained from the biometrics characteristic. So, remember their feature vector is also of is not of same size. For example, that today you have come and you have given the data you may give the data which contains n minutiae points. Tomorrow you come again and if I take your data it need not be that n minutiae points, it may be m minutiae points, an intersection between n and m may be k .

What it means, that today whatever minutiae points you have obtained from a fingerprints and what data you obtained yesterday, these two are not giving the same minutiae points. May be some of them are yesterday they could not that your device could not get, today it got it and another one is that today you did not get, but yesterday image could provide that minutiae points.

Also, in both the cases there is a possibility of false minutiae points what it means that, as I told you that there is break, this break will give you the false minutiae points. So,

there are several issues one is the false minutiae points, true minutiae points. In the case of false minutiae points can occur in both the place, both the times and in the case of true minutiae points also it may so happen today I have got some true minutiae points, but yesterday it did not get and the today - again today I did not get yesterday I got it that may be a possible case.

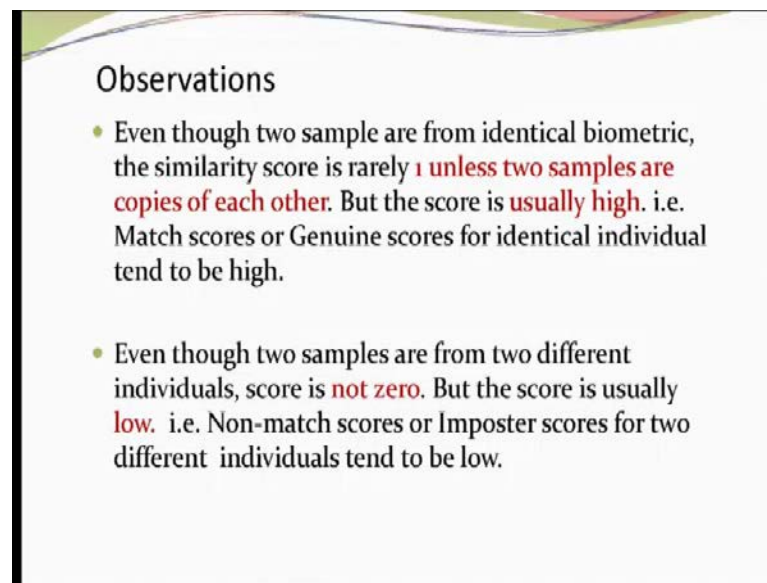
So, by based on that s is the similarity measure between the two samples and when similarity measure the word is similarity is important here, then when s is greater than some threshold value or greater than some threshold value, you tell your null hypotheses is true. Null hypothesis is the two samples are matched otherwise, you will be telling that your alternative hypotheses true, that means the two samples did not match.

Now, sometimes taking such a decision becomes very difficult, it very difficult because based on the matching score, you are taking that match or do not match. Now, there is no scope for you to tell closely match this is very say I tell 50 out of 80 if it is there, then you tell match. Now, I suppose I got 40 49 49 of 80 can you tell matched or not match, you will be telling because you have to tell that, this is our threshold value if it is at least 50 out of 80 match, then only you tell except. So, 49 out of 80 it is less than that you will be telling him that throw out, telling that no it is not matched.

But in reality, you will be telling closely match or sometimes that it is a borderline case, you tell you that I do not know it is a borderline case, but that type of decision is not allowed and this is known as decision making without any exception handling. And this type of reliability of your score is dependent on this type of parameters, that whatever score you have obtained s value you obtain they are dependent on these three parameters. These parameters are all because of variation in data acquisition as I told you that today is bad day for me, so I my biometrics data is bad but, it may so happen.

So, variation is in data acquisition, variation is in sensor, variation in data representations and these variabilities are very high compared to variability in authentication method like password or key and so on. Because, in the case of password, you can make some mistake in entering the data, nothing more than that. But here it is not with you, you have come, you are genuine person, you have given your data, but because of this type of parameters you are facing problem.

(Refer Slide Time: 27:54)



Observations

- Even though two samples are from identical biometric, the similarity score is rarely 1 **unless two samples are copies of each other**. But the score is **usually high**. i.e. Match scores or Genuine scores for identical individuals tend to be high.
- Even though two samples are from two different individuals, score is **not zero**. But the score is **usually low**. i.e. Non-match scores or Imposter scores for two different individuals tend to be low.

So, the variation is very high compare to the other one. Now, I am person I have come the date to give you the data twice, even I get the data within an interval of 5 minutes, you observed that ideal environmental that my matching score should have been one. If it is normalized like between 0 and one, I will tell the score is one because same person within 5 minutes you have given the data that my feature should be such that they are always matched - matched all features should be matched I should get my matches score one. But it is not the case it is only when you will get the one, when the same photograph you have used same image, you have used twice for matching.

However, it can be told that yes that matches score will be near one or is very large as time distance between the two image is small. So, generally for the genuine case matching score will be for the similarity case, the matches score will be large or nearer to one and so on but it will not be one.

Suppose, you get one that may 100 percent match, you can immediately draw a conclusion that the second image is the copy of the first image. Similarly, that if the two samples are acquired from the two different persons, then obviously the matching score should be 0. The two minutiae points of the two samples, if I try to match it should not match at any case. But again this is not true, there it can there will be out of 80, some minutiae points you will find they are matched, but then number is very less compared to the minutiae points of the same subjects.

(Refer Slide Time: 30:15)

Decision

- Accept H_0 if score s is greater than T when in fact H_1 is true.

It means, in case of False Match, an imposter has somehow generated a high $s (>T)$ non-match score i.e. a subject has impersonated another given subject.

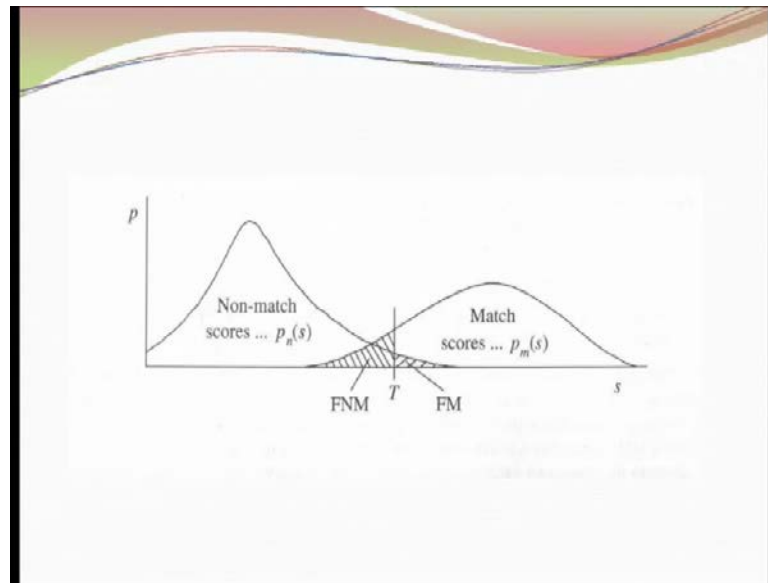
- Accept H_1 if score s is not greater than T when in fact H_0 is true.

In means, in case of False Non-Match, a genuine subject has caused a low match score $s (!>T)$ i.e. a subject has wrongly rejected

So, it is not 0 but it is low. Now, you tell that my null hypotheses is accepted what, when you are telling that if my score is greater than T . Otherwise T is a threshold, otherwise you tell no my H line is accepted, what it means in case of false match - false match imposter has come with some idea and he has given to you an accidentally the matching score has gone up, say in the case of signature I forget you know make several trials, after that he comes and he gives us the signature of one person and then there is a possibility the matching score between the two signature characteristics are is very high. So, imposter is lucky and I have to make it throw to tell you that H_0 is, H_0 is accepted.

Now, H_1 is accepted what it happens, in that case s is not a greater than T and as a result you will be accepting H_1 not H_2 0 here it is happening like this that my day is not good day and I have come from the field and I have to immediately, I m giving you the data and accidentally that all the biometrics features, I am not able to extract from your characteristics. So, you will not be able to draw any conclusion and because s is very small and so you will be drawing that no, it is not matched.

(Refer Slide Time: 32:02)



So, the this diagram tells you whatever I covered till now, based on this there are two graphs. One is the graph for genuine persons, another is graph for imposters. Now, what happens that if I had a sample size of n or n peoples biodata I have taken, their feature vectors I obtained and (θ) if I put the matching score what will happen, see they are genuine persons, their values will be matches score will be in higher side see here I have taken say it is minus infinity to plus infinity say.

So, the matches score will be in higher - it will be higher as it old in the beginning that or few slide back, that if that features from the two subject - same subject, features of the same subjects are match then score will be high. So, the score will be this is a scores, this is the frequency so there will be people - more number of people, more number of people will have the score, this score is very high.

So, as you know for large value of n or number of population, it obviously the normal distributions so this will obviously will at type of thing so very few people, will have this type of scope or very few people have the infinity - score of infinity or say, but at the same time because as I told you that, two samples will not give you one value, one or infinity but most of them will be here, most of them will be here. Similarly is the case with the imposters, when the case of imposter the scores will be more smaller, compare to the case with genuine. And you will find that, this peak will be much closer to this side, much closer to this side because most of the cases it will fail, but very few of them will have the high score.

So this region, so this will be a graph for imposters, this will be graph for the genuine and there will be a some overlap because of pattern matching as I told you. So, if I consider that this is a threshold, so all the scores or all the persons whose score, matching score is greater than T, you will be accepting that means these people even though, they are imposters they will be accepted and that is known as false match. So, the total number of cases out of this occurrences you will be your false acceptance rate and this people, because of this threshold they will not be accepted wrongly, this is known as false non match.

(Refer Slide Time: 35:47)

Error Rates

- False Match Rate and False Non-match Rate are frequencies at which False Match and False Non-match occur respectively.

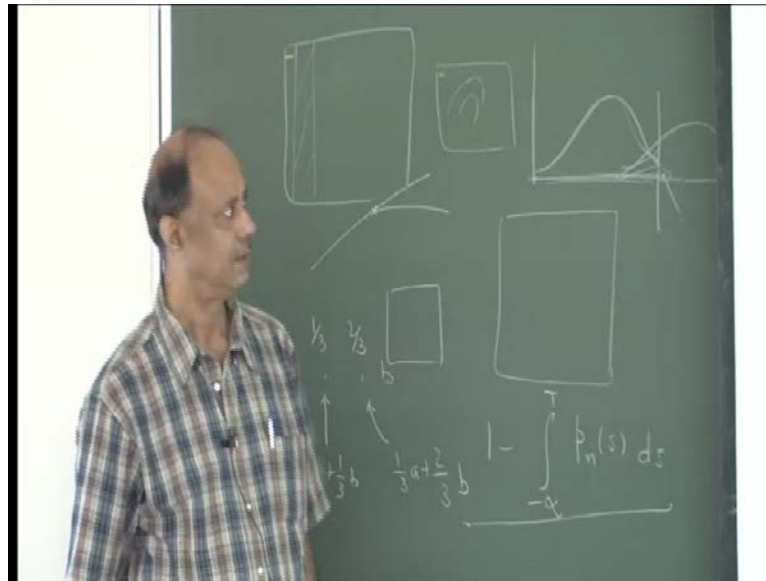
$$FMR(T) = 1 - \int_{-\infty}^{s \leq T} p_n(s) ds$$

$$FNMR(T) = \int_{-\infty}^{s \leq T} p_m(s) ds$$

- For biometric applications, match score distribution and non-match score distribution **always overlap**
- It is not possible to select a T such that **FMR = 0 and FNMR = 0**
- T should be selected in such a way that the system operates in **optimal fashion**. In this figure, T is selected so that **FMR is less than FNMR**.

So, if it is the case can I define my false match rate is 1 minus minus infinity to T polarity density function of that is minus infinity to T 1 minus this is the one, this is the 1 1 minus minus infinity to T this is graph.

(Refer Slide Time: 36:18)

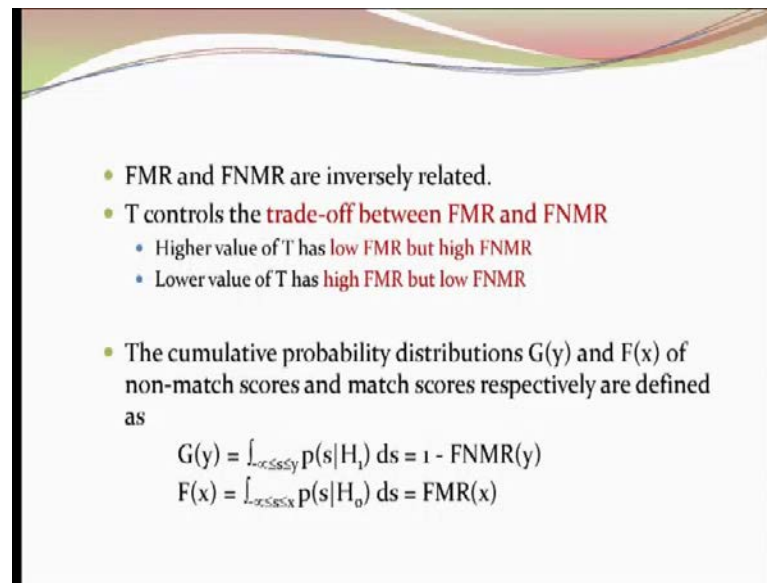


This is 1 minus, this false match rate yes false match rate is what this is false match rate, that means 1 minus, this is what 1 minus this part, so this is your false match and the false non match rate is minus of this part, minus infinity to this of this curve.

So again, if you see the distribution you observe that they will be overlapped, you will not get a situation where both FMR and FNMR they are 0 you will not get, because there are been overlapped. Because this is a Para recognition problem, so only thing is that you have to select that T based on that you have tell whether false match rate or what should be my false match rate or what should be my false non match rate. That T is important, based on the application because in some application I may not lie to allow any suspected people, so what I will be doing that there should not be any false match.

Now, I want to make it false match 0, I do not care what should be the false non match. Similarly, there are some application where I want to make false non match is 0, but I do not care about the false match. So, but you have to decide your T in such a way that this my condition satisfies.

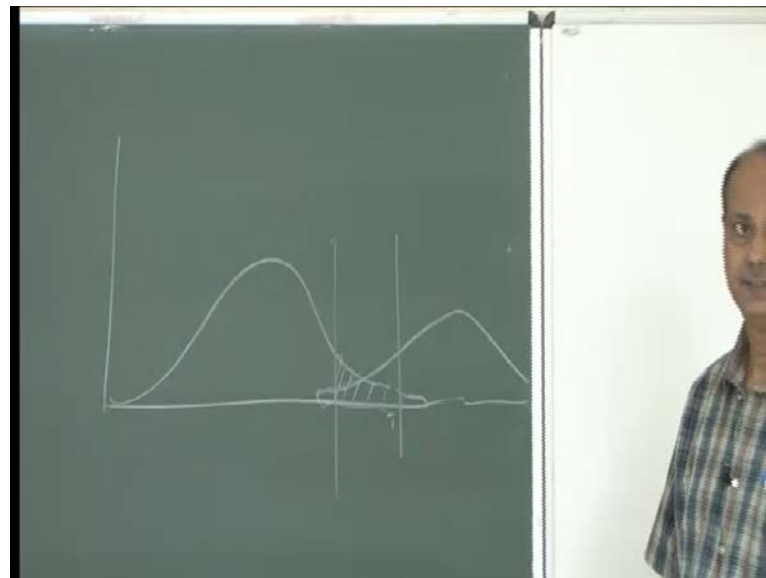
(Refer Slide Time: 38:29)



- FMR and FNMR are inversely related.
- T controls the **trade-off between FMR and FNMR**
 - Higher value of T has **low FMR but high FNMR**
 - Lower value of T has **high FMR but low FNMR**
- The cumulative probability distributions $G(y)$ and $F(x)$ of non-match scores and match scores respectively are defined as
$$G(y) = \int_{-\infty \leq s \leq y} p(s|H_1) ds = 1 - \text{FNMR}(y)$$
$$F(x) = \int_{-\infty \leq s \leq x} p(s|H_0) ds = \text{FMR}(x)$$

What you observe that these two are inversely related, so T controls this trade of relationship between these two, if T is high because it is a similarity measure.

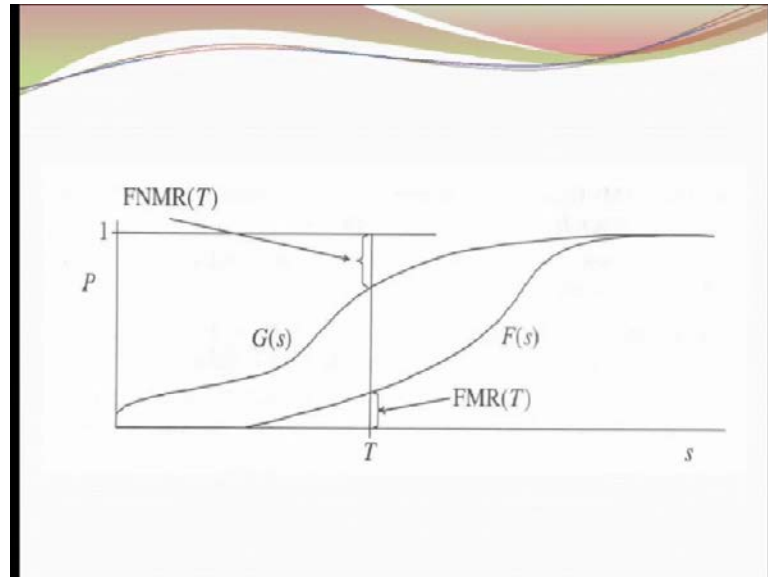
(Refer Slide Time: 38:57)



If threshold is high, if T is high then what happens this is becoming less, but false non match is becoming high. But if T is small, T is here in the case of similarity measure false non match is less, but false match rate is high. So, you have to decide your FMR and FNMR based on your T, now there are three factors are coming in your mind. One is T, another one is false acceptance and another one is false rejection. And these there parameters are you know, if I know one of them I know other two, if I know the threshold I know what is FM and what is FNMR, if I know the false acceptance I know

what is T and what is false rejection, if I know false rejection I know the other two, so they are related.

(Refer Slide Time: 40:27)



Now, the cumulative density function also you can define and this if you draw in the diagrammatically, you can see the diagram of these two will be like this is a simple diagram.

(Refer Slide Time: 40:39)

- Given the score densities, one may like to compute the probability of a False Match and of False Non-Match for some given similarity or dissimilarity score s .
- But the **probability cannot be computed** exactly but can only be **estimated** using test databases.
- True value of **False Match Rate** is **very hard** to estimate [even though one has large database but there is usually a little or no data available about these forged or impersonated biometrics.].

So, given that score densities we like to know FAR and FRR I know the score densities, I want to know exactly what is my FAR and FRR, but unfortunately you cannot do that, you cannot get exact value of FAR and FRR, because you do not have their total test

data. So, it cannot be computed exactly but you can give the estimated data based on your test cases. Now, this is also very hard to obtain the false match rate, because to obtain the false match rate you need to know or you need to get some imposters. So, you know if I tell you that, can you copy his signature you will start thinking that it is my teacher is telling, because once I do it I may be in problem then he has the habit of forging others signature.

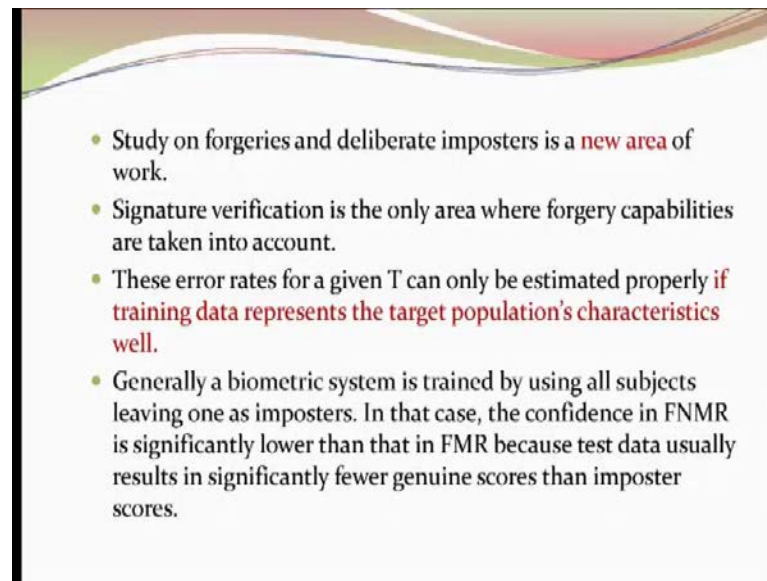
So I can give you one example, you know one day I kept the key of my door inside and then I latch the door accidentally, obviously nobody will do it, accidentally because I kept my key and I latch the door. Now, I know where is the key because you know everybody keeps his key in appropriate in some positions, I know the key that it was evening dark then I called somebody, that can you help me to take out the key. Then I explain that, it is there and I need a bamboo where he will put one iron like this and he will bring it, because there is a key ring is a simple process and then I told that, I have a torch light also, I will provide everything just you keep bring it.

Then, he saw this position yes the key is lying there, I gave him the small bamboo stick type things, iron things and everything but he suddenly told that, no; I will not do it, I do not know why it is simple, today if I do it then tomorrow you will tell that I have the habit of doing this that I know how to take out the key from that area and other things. So, at that time you know in the campus was not safe. Nowadays, I do not know but at that time campus was not safe every alternative day, there is some theft cases stealing small small items. So, he was afraid of this is so, similarly if I tell you that can you sign on the check leaf, I will give you my signature and copy it I want to test whether some bank people can catch you or not. Then also you will think no I will not do it, because you will think that what is the reaction afterwards, I do not know. So, getting imposter is not that simple so as you do not have the imposter, how can you find out the false match rate. See what happens what we were doing correctly is a very wrong method, but we are doing it, we are showing the false match rate is this because suppose I have the face database, I have the face database and I want to find out whether what is the false match rate what is FAR.

What I do, I take your photograph and I matched with everybody else and again I take his photographs and matched again everybody else, that means one leaving one and I am matching with everybody else and then how many times it matched, based on that we obtain that but it is does not mean make sense in reality, how can I take your photograph

and match with him does not make any sense, but we are doing T to draw the conclusion of FAR. But in reality, that is not the correct thing, in reality is that you have come or somebody else has come with your face and whether it has been matched or not. That should be the thing, that I have put some mask on my face which is similar to your face and whether I have been caught or not that will be your false match.

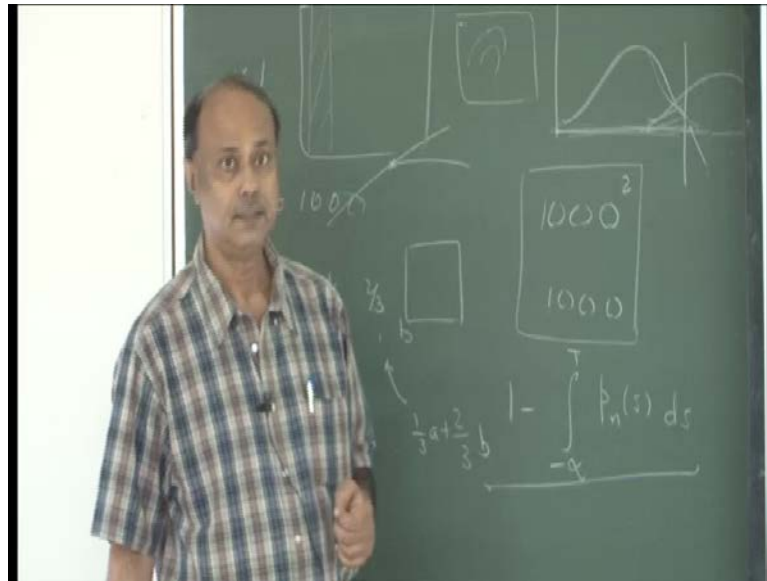
(Refer Slide Time: 45:28)



But the people, there is a strong group of people who does the or who works to forge the signature, that is possible. If you, if I try several times I am sure I will be able to copy your signature, copy your working style, copy your voice this may be possible, but it is not possible to copy my face, you cannot copy my face characteristics. So, the people are working on signature forging and it is a very good field of work also, it is a good field of research. By seeing your signature, can I copy your signature is not that easy also but it is a work is going on that and this error rates for given T can only be estimated properly.

If this represent the target population characteristics that means if I have the database and if I understand carefully that behavior of each characteristics in the database and then you design your system. What it means that, suppose I have a population of size n, which is not available with me but, I have a training set of size small n, this small n size people they are covering the whole population characteristics, that they are from different background, different age groups, different genders and so all characteristics are in built we in this training set, then possibly this error rates that you can obtain properly.

(Refer Slide Time: 47:52)

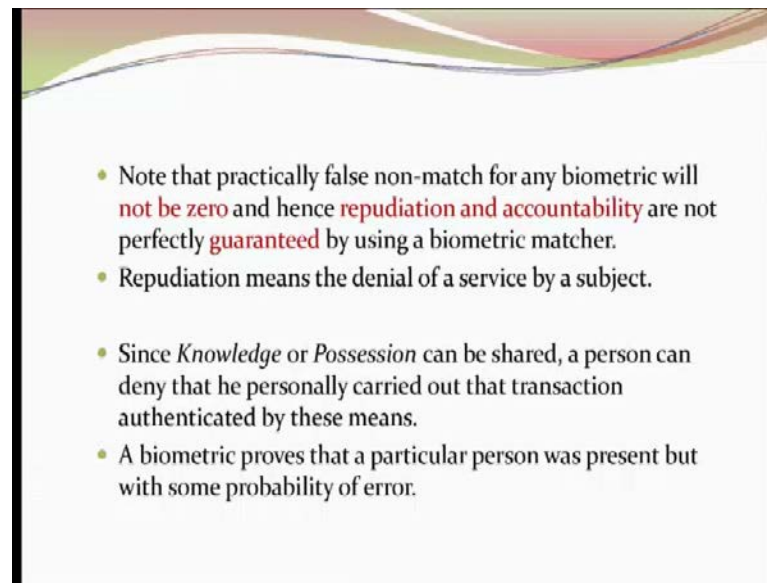


Now, since we are using one leave out method while computing false match rate one leave out means suppose, I have a database n is 1000, small n is 1000 for genuine image. What I do, I take one person and I am checking the data with this person, that designated person that the person with identity one is compared with the database of the person say with identity one only second and database second, third with data base third so genuine match you can measure and this is 100 such cases, a 1000 such cases because I have a 1000 database size.

Now, in the case of imposter one leave out because I will be leaving out myself so, 999 I will be compared with remaining 999 people and there are 1000 such people. So, 1000 into 999 size that means 1000 to the power 2 so in the case of estimating false acceptance rate for a database size of 1000, I have the result on 1000 square approximately there is a 999 into 1000.

Whereas, in the case of genuine, this is genuine because false non match rate I have the size is only with 10,000 or 1000 because my database size is 1000. So, the genuine case I am for a genuine that a false non match rate is estimated based on 1000 data and false match has been estimated on one 1000 square data. So, that is the thing we are, to want to write here the confidence on false non match rate, because size is very less is less compare to the false confidence on false match rate, because false match rates size is 1000 square in our database.

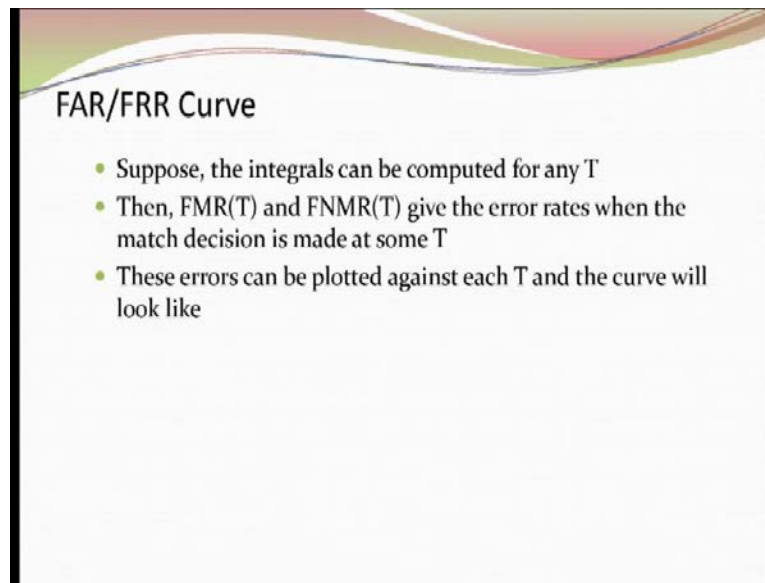
(Refer Slide Time: 50:06)



So, you observe that false non match cannot be 0 because if you make it 0, then false match rate also will be very high there it will be near 0, you can tell but this is not 0. If it is not 0, then I am facing two more problems one is reputation, another one is accountability. That means I will be denying some genuine person to use the system, because non match rate is not 0 that means there will be some people, genuine people for them I will tell no it is not matched and he will be unhappy. So, genuine person will not be allowed to use the system.

So, in the case of knowledge or possession this you can be shared so a person can deny, that I have not used it or because I have the key or I have the password and I have done something wrong. Later on, you can tell no I have not used, may be somebody has taken my password or sir my password has been stolen this happens, while you see that in some activity. Whenever we catch some students in the computer center's activity, sir I do not know who has taken my password and all passwords are freely available. So, he can always tell that yes I have not done this work he can always deny, but this is not the case with the biometrics but again you cannot tell that, no there is no error, there may be some error because of pattern rate, as I told because of that accidentally he has been matched.

(Refer Slide Time: 51:59)

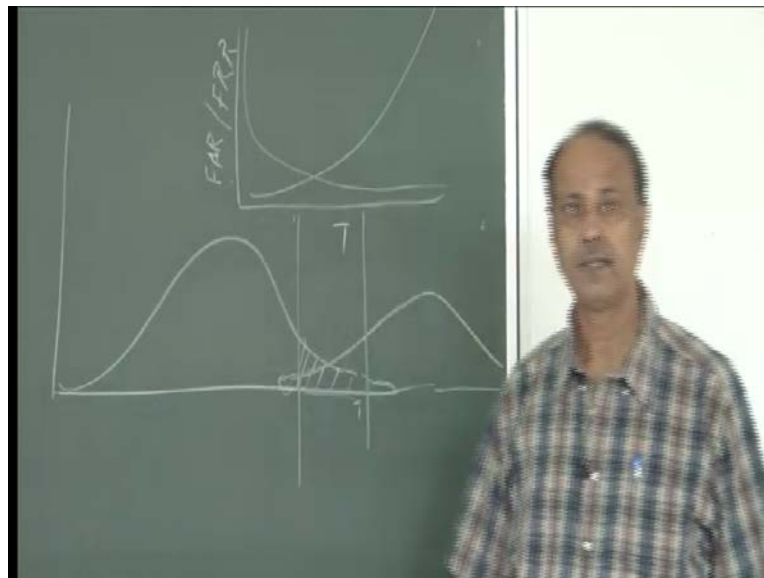


FAR/FRR Curve

- Suppose, the integrals can be computed for any T
- Then, $FMR(T)$ and $FNMR(T)$ give the error rates when the match decision is made at some T
- These errors can be plotted against each T and the curve will look like

Now, you have the world false acceptance rate and false rejection rate or false match rate and false non match rate and you have also a T the threshold value.

(Refer Slide Time: 52:24)



Now, if I draw I can have the two graphs here it can be false acceptance rate or false rejection rate and here I have T threshold. So, what you have you have I am drawing a graph, now threshold you have and you have false acceptance rate and false rejection rate. Now, in the case of false acceptance rate it is the graph I am talking about with respect to the similarity match. Now, if T is small, T is small what will happen false acceptance rate will be high and if T is large, so graph will be like this right and similarly, the another one will be like this. So, this is known as FAR verses FRR graph.

(Refer Slide Time: 53:27)

Receiver Operating Curve

- Suppose, the integrals can be computed for any T
- Then, $FMR(T)$ and $FNMR(T)$ give the error rates when the match decision is made at some T
- These errors can be plotted against each other as a two dimensional curve

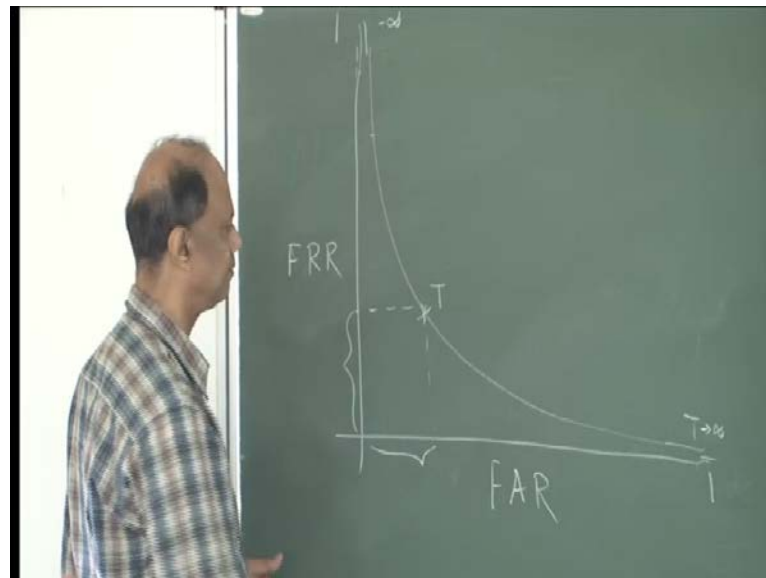
$$ROC(T) = (FMR(T), FNMR(T))$$

- That is, FMR and $FNMR$ behavior is expressed in terms of Receiver Operating Characteristic (ROC) curve
- FMR and $FNMR$, as functions of T , are mapped as

$$ROC(T) = (FMR(T), FNMR(T)) \rightarrow \begin{cases} (1,0) \text{ as } T \rightarrow -\infty \\ (0,1) \text{ as } T \rightarrow \infty \end{cases}$$

Now, again this FAR and FRR graph also can be represented in different way, which is known as receiver operating characteristic graph or curve.

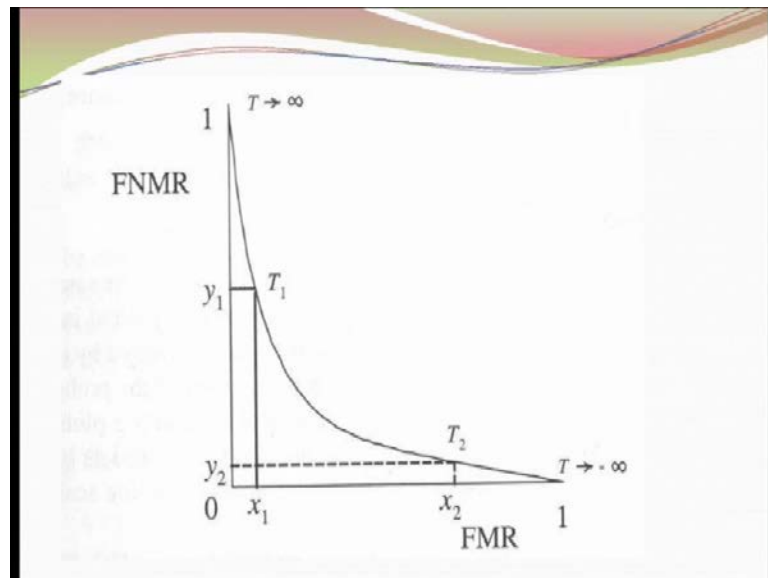
(Refer Slide Time: 53:42)



So, this ROC gives you that at different threshold value as I told you that, if I know the threshold value I know what is FAR and what is FRR. So, I can have similar type of graph that when threshold is at threshold some minus infinity, you will be getting FRR is this, FRR is this and then this point what is the threshold at any some particular threshold value FRR is this and FAR is this you can get all the term, yes or no. This graph will be there so this is FAR verses FRR graph, this is known as ROC curve.

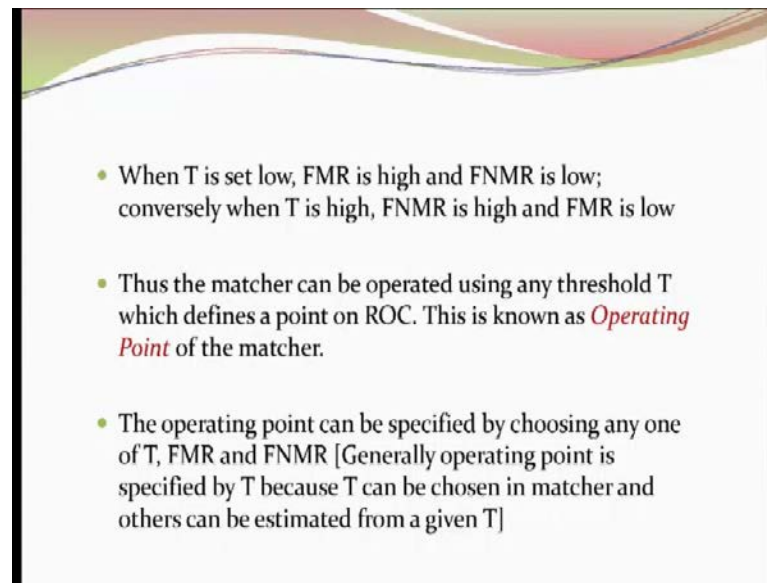
Now, this ROC curve is false acceptance rate is near one, it cannot be one but, this will be 0, near 0 false rejection rate will be 0 and this is one and this is one but it is near 0 for FAR, FAR 0 but FRR is one and the T is here minus infinity and here T is plus infinity and this for a particular T for a particular T FAR is this FAR is this, and FRR is this.

(Refer Slide Time: 55:39)



So, suppose I want small false acceptance rate or less acceptance rate, but so less acceptance say, $x < 1$ then you know what is my false rejection rate and what is my threshold you can easily obtain. See the graph, I started with the three parameters T FAR and FRR now you observe that one is redundant, T is not required. So, we what you have done the graph we obtain FAR and FRR irrespective of my T, it do not need T is your thing that designer has to decide what should be my T. So that, I get this FAR and obviously I want one near 0 and corresponding value of other one.

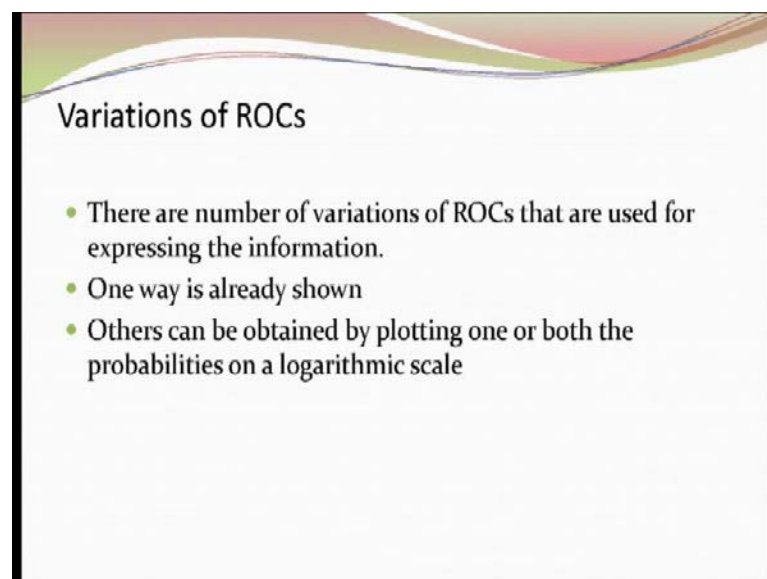
(Refer Slide Time: 56:35)



- When T is set low, FMR is high and FNMR is low; conversely when T is high, FNMR is high and FMR is low
- Thus the matcher can be operated using any threshold T which defines a point on ROC. This is known as *Operating Point* of the matcher.
- The operating point can be specified by choosing any one of T, FMR and FNMR [Generally operating point is specified by T because T can be chosen in matcher and others can be estimated from a given T]

So, this selection of T FAR or FRR this is termed as operating point. This point is used to decide that what should be my system, what is my performance of my system everything is dependent on this operating point, I think these are things I already covered.

(Refer Slide Time: 57:02)



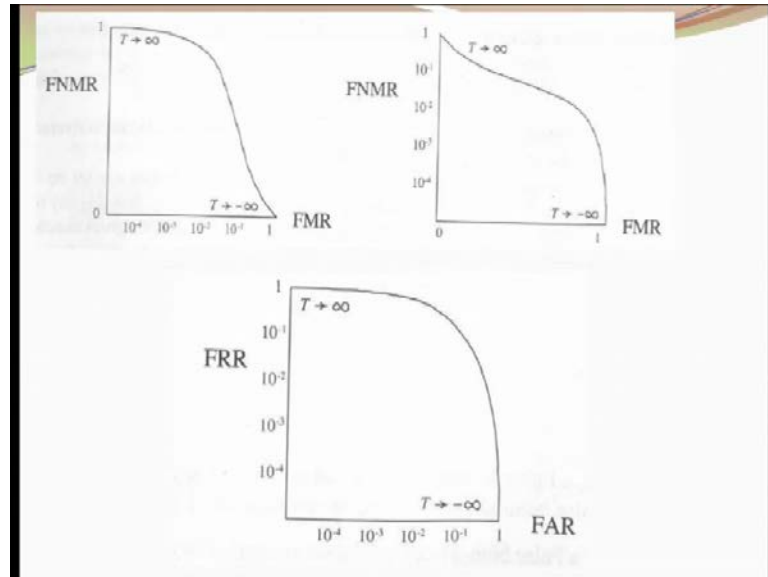
Variations of ROCs

- There are number of variations of ROCs that are used for expressing the information.
- One way is already shown
- Others can be obtained by plotting one or both the probabilities on a logarithmic scale

Now, this ROC curves one ROC curve I have shown right and this is not only one type, way you can represent you can have several ways this graph because FAR is what false acceptance, that means what there are people some imposter they have entered into your system and one leave out method it is of the order of square. Suppose, I have the

database of 10,000 one leave out that number of persons will be here is 10,000 square and this side is your on the 10,000.

(Refer Slide Time: 57:59)



So, what we do sometimes the log graph we introduce so log graph can be in with respect to FAR may be with respect to FRR may be in both the side depending upon the database size.

(Refer Slide Time: 58:15)

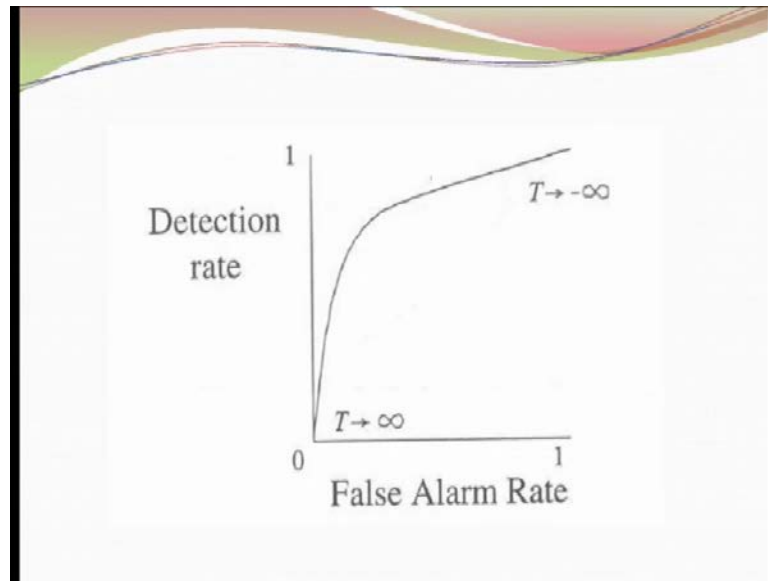
- Sometime one plots Correct Match Rate [i.e. $1 - \text{FNMR}$] against FMR. This is called **Detection Error Trade-off [DET]** Curve.
- Along y-axis, we have $[1 - \text{FNMR}]$ which is correct detection rate and along x-axis, we put FMR [False Alarm Rate]
- Detection Rate goes to 1 when False Alarm Rate goes to 1

So, the another graph which is known as correct detection error trade off graph, detection error trade off graph which is known as DET. Here, I have correct match rate, correct match rate means 1 minus false non match rate. False non match 1 minus false rate

match rate, false non match rate will give you the correct match rate against false match rate, if I drop this graph that graph is known as DET. So, here instead of telling that false match rate, we use the term false alarm rate.

Now, obviously the detection error rate or the DET error trade off will be lying between 1 and 0, this will be one when false alarm rate is one, then the graph will look like.

(Refer Slide Time: 59:18)



Now, if I give you matching score if I give you say ten cross time matching score table, can you draw FAR versus FRR or some suitable threshold. Obviously, you have to decide what should be my threshold, can you draw ROC curve, can you draw DET curve, can you think about it and if you can cannot then let me know that means in that case, I have to explain little further and if it a no sir, it is possible then I will not explain.

What I have, I have a table which gives you the matching score value. Matching score means obviously you will be telling sir is it a similarity match or dissimilarity match. Suppose, it is similarity match then it is lying between 0 and 1, that is also given now what I am telling you that I want ROC curve.

So, for different threshold value you have to first find out, what is your false acceptance rate and false rejection rate, you can find out, you have a table and based on that you decide what should be the threshold value. Different threshold value you take first, you take threshold value 0 then you take 0.1 then you take 0.2 0.3 0.4 0.5 like that you can take or you could have taken 0.1 0.11 0.12 0.13 like that also you can take whatever threshold you want to decide or if you see that by the table, you see that all the points are

lying between 0.4 and 0.5 then why to take the select the other things you select 0.4 0.41 0.42 0.43 0.5 all those 10 intervals also you can take whatever, threshold you decide for that each of them you can obtain the FAR and FRR. Then you can draw the graph FAR verse FAR graph and FRR graph then you can obtain ROC and finally, you can obtain also DET it is not difficult, is it difficult.

Now, if I tell you that difficult will be coming little for when I tell you that can I draw this, if I give you the two matching score graph diagram or table they are of different, they are not normalized means one table has the matching score lying between 0 and one other one is may be minus 5 to minus 7 plus 7 and so on. So you have to normalize them.

Problem will be little difficult further, if I tell that one is similarity based on similarity score and another is dissimilarity score. So, you have to make it after normalization 1 minus that you have to take to make it same platform, similarity platform or dissimilarity platform.