2003

The application of face recognitions on evaluating sensors

Sui-Yu Wang
Lehigh University

Wang, Sui-Yu

The Application of Face Recognitions on Evaluating Sensors

January 2004

The Application of Face Recognitions on Evaluating Sensors

By

Sui-Yu Wang

A Thesis

Presented to the Graduate and Research Committee

of Lehigh University

in Candidacy for the Degree of

Master of Science

In

Department of Computer Science and Engineering

Lehigh University

November 18 2003

This thesis is accepted and approved in partial fulfillment of the Requirements for the Master of Science.

12/4/03

Date

Thesis Advisor

Chairperson of the Department

Acknowledgements

I would like to express my sincere gratitude to my beloved parents, for without their support and encouragement I would not have went to graduate school in the first place. Special thanks are due to my advisor, Terrance Boult, who gave me directions and guidance that made this thesis possible. I also appreciate the strength my sister and my cousin gave me when I was down. I would also like to thank the friends and lab-mates who I benefit a lot from the discussions with them.

Contents

List of figures

Abstract

The success of a face recognition algorithm depends on its robustness with respect to various factors such as the lighting condition, the departure of the face from fully frontal, the skin color, ...etc. The performance of the sensors used to capture the test image also plays an important role in acquiring images for basic feature extraction. While the ability 7of the sensors might affect the image by the discrepancy in pixel level (not necessary a relationship of inferior and superior), no previous research concerning how and to what degree this disparity could affect the results of the face recognition algorithms has been conducted. Using the results of the same face recognition algorithm performing on different sets of data that have exact image context under the same condition except taken by different sensors gives results of evaluating sensors directly coming from the impact of the quality of sensors on the algorithms. Similarly, evaluating the same sensor under different outdoor weather gives the impact of the weather directly on performance of the algorithms. In this paper we conducted evaluations that clarify the effect of the weather on sensors and the difference between sensors and proved that these factors can result in considerable difference in the performance of the algorithm. In addition to the general results concerning sensor's ability, we also explore various ways to use limited data to do evaluation that gives us additional information about how various factors are affected by this difference. The empirical distributions used for statistical tests are formed by using the Central Limit theorem, which states the limiting property of the summation despite the different true distribution of the variable used. Various formation

of the Central Limit Theorem and the criteria that make them work will be introduced throughout this thesis. Different formation and the influence of the factors will be explored in different experiments. The results of this paper suggest any experiment undergoing sensor change should be cautious about the effect of the difference in sensor to ensure the consistency and stability of their results.

Chapter 1. Introduction

Face recognition has been one of the most studied fields in computer vision. Various efficient algorithms have been proposed to adopt different inferior conditions [1, 5, 26]. Currently there are four major categories of algorithms following different principles [9, 21]: PCA [2], LDA [27], Bayesian classifier [15], and graph matching [10]. The PCA approach tries to find the most important features and ignore others to reduce the dimension and calculation needed. After extracting the features, several "eigenfaces" are serving as the basic unit vector. Test images are represented as the combination of these basic vectors. Recognition takes place when the test image's vectors are most close to that of the recognized image. The most recent LDA method developed by University of Maryland is actually a combination of PCA and LDA, it uses PCA to obtain some major feature then use the linear discriminate method to make the decision [27]. The concept of LDA is similar to that of PCA, but the meaning is in a complemented way. While PCA tries find the most important feature, the small difference that distinguish the two character "O" and "Q" might be ignored[book].While for LDA, the approach is to sum the features together. Thus the small tail in "Q" will make a difference between them.

The Bayesian approach is developed by MIT [15]. The Bayesian classifier, unlike those previously stated, uses a probability model rather than some distance measurement to predict the probability of the face being recognized. Instead of using

eigenfaces, they use the intensity curve of the face for fitting and predicting. The graph-matching face recognition was developed by USC[10]. This method turn the fiducial points of face (eg, eyes, mouth,...) into graph, then turn the face recognition into graph matching problems.

The FERET database construct first by University of Maryland consisting of test images and algorithms gives a standardized evaluation for face recognition. This solves the problem of researchers claiming the performance of their algorithm with possibly biased measurement without fair criteria. Face recognition applications have improved to the extent that it can be used in real-world applications. Face recognition can robustly recognize people with hair style change or some other disguise techniques since it is mainly based on the structure of the face instead of some other appearance that people used to recognize face. Applications range from custom check to personal identification. Some commercial applications has been made in the market like the Visionic[6].

Despite all the progress and effort researchers made in the literature, there still exist some experimental problem that hasn't been addressed. Since face recognition lies in the feature extraction, the source of the feature, the image quality, or equivalently, the sensor's quality, comes into play. Under practical situations the image taking process might have to undergo sensor change [14]. Even in a single sensor, the response curve of each pixel may not be totally identical. Without taking the effect of sensor change into consideration before applying target experiment might cause inconsistent results. The role of sensor evaluation become even more important while the application of face

recognition has moved on the real world application, where the deploy of the camera is inevitably in outdoors, and unlike indoor applications, various factors like lighting conditions can not be reproduced, which might worsen the consistency of images taken by different sensor due to sensor's nonlinear response to lighting condition. Even a single sensor can have a different result for fixed scene and fixed under the same lighting condition. The difference in temperature may cause the sensor to operate in different operating environment thus have a different response curve.

Since all recognitions lies on extracting basic features from the images, less-sharp images reduce the information available for recognition. Not only the blurring of the image could cause problem. Sensors from different models or producers might make the sensors response to lights from different wavelength differently. This might put different emphasis on different features. This difference in the image might cause the algorithm to perform differently for the same target and scene. Although sensor's response to various situations can be modeled by the user manual as done in [4], how and to what degree does these differences effect of recognition has never been explored/remain unknown

Given a set of images taken by some sensor, the distribution of the recognition results represents the algorithm's response to the various parameters of the training set: the condition under which the images were taken, and the sensor's performance. While the true distribution of what the algorithms should response to a perfect camera is unknown, and we cannot quantify the how performance of the algorithm has departure

from the true performance it should have for an image taken by some perfect camera. By comparing the sets of results taken by different cameras by statistically justified method we get to know the potential difference among the sensors might have from those distributions of results.

The results of the recognition for images taken from some sensor can be transformed into empirical distributions identifying the performance of that sensor regarding particular conditions (if there is any) like lighting conditions under which the images were taken. We address two important issues concerning the formation of the distributions. First, the distribution should be able to represent the sensor's ability as a whole. The sensors may perform differently on different class of images. On some special cases like the target's wearing glasses, blurred images may be better for recognition than clear images that separate the frame of the glasses from the eye. The overall distribution of the sensor should be able to show the superior/inferior of the sensor despite these special cases. Second, to make the distribution represent the sensor's characteristic more accurately, we should exploit all knowledge we have about the images to reduce the variances. While some factors like a particular gesture may tend to be more sensitive to the changes in image quality than others, we look into these factors by using them as the basis in the formation of our distributions. By comparing distributions based on different factors we get to further investigate the sensitivity of different factors to the sensor change.

Two cameras of the same factory quality are placed at different distances from the target to simulate different sensor quality. Our experiments are conducted outdoors to simulate the environment the researchers are most interested to solve. Several different factors can account for the cause of different sensor response for a single sensor response. We try to eliminate this intra-sensor difference by collecting sets of data under similar conditions. Factors can be put consideration includes the visibility of the scene, the wind speed when the images are collected, the temperature, and the sky condition, the humidity, ...etc. If we specify the factors in a strict way, the data belong to each category might not be enough for us to conduct statistical procedures and tests. In our current implementation we only specify the sky condition roughly into three categories, clear, partly cloudy, and mostly cloudy. This way it would allow a large pool of data to be collected. This pool of data includes different targets (people) under different pose of that target. We take totally 256 people and four poses of each person as our raw data. Some algorithms are then used to recognize these faces. The metric used to characterize the performance of the sensors follows that in [14, 19]. The algorithms took two sets of images. One is the training set, which gives the algorithms a standard of what do identify. The other is the testing images, which is the target to be identified and find a match in the training set. These two sets of images are exclusive. The set of test images different from the training images is fed to the face recognition algorithm. The algorithm can compute some score saying how similar the test image is to one of the images in the training set. To do classification, the algorithm has to

compute the scores for the test image and all the images in the training set. A success of the recognition should rank the probability (score) of the correct face being the highest among all. Failure may result from either miss-classification or failure in detecting the faces at all. The lower the probability the algorithm produces for the correct class, the lower score it receives in that test. Different levels of failure can either receive different scores for how serious the algorithm had done wrong or it is simply a mistake no matter how good or bad this mistake is. Details in defining the test and the metrics used will be stated in next section.

Simple looking into the test result of each image does not give us too much information other than how the algorithm works under this specific person, the pose of that person, the condition like lighting of weather in the specific scene. We tried to group similar data into some distribution. The distribution of the results for each camera represents the performance of each sensor for the specification of that group of data. The pool of data collected may bear several kinds of variances thus makes comparing the mean of each camera meaningless. By combining data in different ways we not only reduce the variances, but also get to investigate how various factors (person, pose, ...etc.) react to the difference in sensor quality. The perfect distribution of the specification of the group of data can never be know, thus we have know way of know how far these data has gone from where it is suppose to be if the camera taking them are perfect. However, the Central Limit Theorem states that the distribution of an average tends to be Gaussian, even when from which the average is computed is decidedly non-

Gaussian [7, 11]. By taking averages of similar data, by similar we mean to eliminate some potential difference like different target, and we can convert the original distribution into something we can use to do statistical tests, we get a Gaussian distribution representing the algorithm's response to that set of testing data.

Although this Gaussian distribution gives us some idea of how the algorithm response to that set of data with its various factors like how the sensor did and the restricted specifics to get this similar data, it still tell us nothing more than that special condition. Although we might want to compare the two cameras on its overall performance, mixing all the data to get this Gaussian distribution bears too many variances, this result still won't tell us too much more than a single mean. We can then specify the factor affecting the performance of the algorithm as the parameter we wish to investigate. After fixing the parameter of interest, we then by combining several different "similar data sets" under the same restriction we put on the factor of interest we get the result to be both stable and general, and reduce the final result to one Gaussian distribution representing the character of that camera with respect to the factor specified. This final combination has a similar meaning of that of linear discriminate methods. It both serves as a dimension and variance reduction. Unlike the linear discriminate, which tries to find the combination that would show the difference of each class most, we treat each class equally since we wish to see the performance of the camera by investigating the grouped behavior of all the class. The definition of class can be viewed as, by specifying the factor in interest, we vary all the parameters except

the one fixed. And the multiple class may consist of data from sets of "complementing" data, meaning the data are related in a way that not only the variances are reduced, the reduction in dimension made the final Gaussian distribution more capable of representing the performance of a particular sensor.

The Central Limit Theorem is not only a theorem. There are series of criteria can be used for various application. The restriction put on the variables in question ranges from independently identically distributed, independently distributed, to dependency between data can be assumed. We will examine the validity of our data before we apply the theorem. After the Gaussian distributions representing the sensor have been constructed, we can then use the t test to see if there is a significant difference between the qualities of the two sensors.

The importance of this evaluation come from that it not only gives a way of evaluating the performance of the sensor, but that the difference tells us directly the result why the researchers wish to conduct the sensor evaluation at all, to see to what degree can the sensor's performance can affect the algorithm's performance. We can then conduct some statistical test from two Gaussian distribution A t test is then applied to the distributions of the two cameras to decide if there is a significant difference in the performance of the two sensors. Our results showed that the difference in the sensor's quality, while might not differentiable to the human eye through the image quality, can indeed affect the performance of the same algorithm to a significant degree. Any

experiment undergoing sensor change should take this factor into consideration before applying further analysis/experiment to avoid inconsistent results.

Chapter 2. System Model

In this section we define the terminology and introduce the mathematical backgrounds and the metrics that will be used throughout this paper. We will first define some basic terminology that has special meaning designated to make the statement of the concepts and experiments easier and unambiguous. The metric defined by other researchers that will be utilized in our experiments will then be stated. Finally, the mathematical principles that will be used to transform the recognition results into empirical distributions designating the performance of the sensor will be introduced with their basic idea and the criteria for them to apply.

First we define the term *experiment* to be the basic unit for testing some hypothesis about which we wish to investigate. In one experiment there can be several *trials* that may have different formation concerning the data. These formations, however, must be constrained in such way that the hypothesis of the experiment can still be tested in each one of them.

2.1 Image classification

In this section we define the terminology and metrics concerning how we define and combine recognition results that will be used in this paper. Some of the definition follows that in [14, 19], the reader can find more details in it.

Each image has certain *specifics*. The *specifics* is a 4-tuple (T, W, E, O) where

1. T is a finite set denoting the time the image is taken,

2. W is a finite set denoting the weather condition,

3. E is a finite set denoting the person in the image, and

4. O is a finite set denoting the pose of the person while the image is taken.

Let $l_p(e_i)$ be the *label* of *i*th element (image) in an image set p, designating the *specifics* of e_i. Each in the database can be uniquely identified by the label of that image and the camera used to take that image.

The set of images G represents a *training set*, or *gallery*, which is used in recognition process as the model that new incoming images should be mapped to. Some image set, p, different from those in G, is called the *probe* or the *test images*, which are to be classified. Each probe has a *property*, which constrains certain specifics of the images in the probe. For any of the tuple X in *specifics*, we write $\boldsymbol{P}(X)$ to be the collection of all subsets of X. $\boldsymbol{P}(X)$ is called the *power set* [5][1] of X. We define a special character D_X, "don't care", to be the largest subset of X, $D_X \in \boldsymbol{P}(X)$[2]. The property of a probe is a 4-tuple (T', W', E', O') where

1. $T' \subseteq \boldsymbol{P}(T)$ is the set of times the images in p were taken,

2. $W' \subseteq \boldsymbol{P}(W)$ is the set of weather conditions,

3. $E' \subseteq \boldsymbol{P}(E)$ is the set of people,

4. $O' \subseteq \boldsymbol{P}(O)$ is the set of poses.

---

[1] The definition of power set can be found in mathematical literatures like [24].
[2] The usage of this definition will appear in later section.

The same image can appear in a certain probe multiple times, that is, it is possible that $l_p(e_i)=l_p(e_j)$, $i \neq j$. We define the *character* of a probe to be a 5-tuple (T',W',E',O',C). C is a finite set denoting the camera used to take the images of the probe. Note here that while the constraint put on the *property* is less strict, that it can be an element of a power set, the camera used in each probe must be *unique.*

A *probe set* P, $P=\{p_1,p_2,\ldots,p_{|P|}\}$, is a set of probes that each probe in the set has a property that is different from others in the set by at most *one* variable other than C in property. For example, all probes in the set has the property (x, a, b, c, d) where a, b, c, and d are constants throughout the set, and x is the one variable that may be different for the probes in the set. Note here that the variable in the property can be one of the four variables, T', W', E', and O'. The camera used for a probe set must be unique. There can be two probes in a probe set having the same property, but the two probes have to be independent, $\exists i, l_{p1}(e_i) \neq l_{p2}(e_i)$. The empirical distributions for the probes in a probe set must be *independent.* Two probes *p1* and *p2* are called *equally representative*[3] if $|p_1|=|p_2|$. Two probes p_a and p_b are called *corresponding* if $l_a(e_i)=l_b(e_i)$, and the characteristics of the two probes differ only in the camera used. Two probe sets are *equally representative* if $|P_1|=|P_2|$ and $\forall p_i \in P_1, p_j \in P_2 : |p_i|=|p_j|$.

Identifying the identity of certain face image is a kind of classification. Each person in the image gallery can be viewed as one class. The ith class in G is denoted as Gi. For a image x in probe p, $x \in p$, and an image $g \in Gi$, the classifier ϕ can compute

[3] Note that the meaning of *representative* here is different from that defined in [14].

some score from the similarity metric $s_\phi(x, g)$, indicating the degree of similarity between these two images with respect to the recognition algorithms used. To classify x, the classifier computes the scores for x and all images in each class in G, $s_\phi(x, g_1)$, $s_\phi(x, g_2), \ldots, s_\phi(x, g_n)$ where $g_i \in G_i$ and n is the number of classes in the gallery. It then outputs a list of classes in diminishing scores. Let the true class of x be *true*(x). Let *rank* be the position the class *true*(x) has in the list. If for some x it has rank 1, it is correctly classified. Otherwise it will have the rank value of 2 through n where n is the number of classes in G. The larger the rank is, the more seriously wrong the classifier did in the classification for x.

2.2 The Central Limit Theorem

The results of individual classifications, the ranks in our experiment, have to be transformed into some empirical distributions that we can comprehend, showing the property of the target sensor, and be used to perform statistical comparison between sensors. While we can get an estimator of the distributions of the given statistics using the indicator function like the This kind of distributions do not give us much information about the performance of the sensors, nor would we be able to compare those coming from different sensors since we have no way to know how different these distributions should be so that we can say two distributions from two sensors are performing different enough. While normal distribution may seem a good way to compare between distributions, taking the mean of the given measurements for

comparison in a certain probe is neither useful nor meaningful since we have no way of verifying if the error term of the true distribution of the ranks should be normal, nor would it likely to be, since the ideal mean should be one while the wrong classification will drive the mean only larger. However, Central Limit Theorem provides us a way to transform seemingly unknown distributions into normal by taking the distribution of the means, as long as we have a sample size large enough [16, 7]. We introduce the definition of Central Limit Theorem and the criteria to make it successful in this section.

2.2.1 Central Limit Theorem on i.i.d case

This version of central limit theorem is the most commonly used [16, 7]. Let $X_1, X_2, \ldots$ be a series of mutually independent, identically distributed random variables with finite mean μ and variances $0 < \sigma^2 < \infty$, then

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} = N(0,1)$$

In other words, the distribution of the means from will have a mean $E(\mu)$ and variance $\sigma/\sqrt{n}$. This theorem possesses a desired property that the limit property does not depend on the particular choice of the original random variable X_n. In other words, the distribution of the means of any i.i.d random variable will be approximately normal.

2.2.2 Triangular Arrays

A triangular array is the set of random variables

$X_{11}, X_{12}, \ldots, X_{1n_1}$

$X_{21}, X_{22}, \ldots\ldots, X_{2n_2}$

$X_{31}, X_{32}, \ldots\ldots\ldots, X_{3n_3}$

$\vdots \quad \vdots \qquad\qquad \vdots$

The array satisfies the following conditions:

1. for each i, the n_i random variables $X_{i1}, X_{i2}, \ldots, X_{im_i}$ in the ith row are mutually independent
2. $E(X_{ij}) = 0$ for all i, j, and
3. $\sum_j EX_{ij}^2 = 1$ for all i.

The random variables in each row are *not* required to be identically distributed. There are no constraints concerning the number of element in each row, n_i. In our experiment, however, all n_i s are equal.

2.2.3 Linderberg's Condition

Although the CLT described above can turn distributions into normal in statistically justified way, the requirement of i.i.d. is not practical enough since taking results of a particular class merely shows the sensors' performances with respect to these specific targets. As mentioned previously, single-class results may be biased due to the properties of the features in that class that might make blurred image better for recognition. These distributions, although can be used to compare in statistically meaningful way, can't show the sensors' performance as a whole. We introduce another version of CLT in this section [11]. This version of Central Limit Theorem relaxes the i.i.d. condition on random variables in section 2.2.1 to the triangular array condition described in the previous section. Let the sum of each row be

$$S_i = \sum_j X_{ij}$$

Suppose that in addition to the triangular array condition, the array satisfies the following condition

$$\forall \varepsilon > 0, \lim_{i \to \infty} \sum_{j=1}^{n_i} E\left[X_{ij}^2 \left(\left|X_{ij} > \varepsilon\right|\right)\right] = 0$$

then

$$\lim_{n \to \infty} \frac{S_i - n\mu}{\sqrt{n}\sigma} = N(0,1).$$

Lindeberg's condition requires that the variances of each row must be small comparing to the total variances of the whole for Central Limit Theorem to apply. This requirement is also viewed as the *homogeneity of the variances*, or *homoscedasticity*

2.2.4 Other Central Limit Theorem

There are other versions of the Central Limit Theorem that does *not* require the variables be *independent*. That is, the sums of non-independent variables. The m-independent central limit theorem [3, 23] states that in the triangular array defined above, the requirement of independent rows can be relaxed to m of them being dependent where the m can be extend to infinity as long as the ratio of m and n, the total number of rows, are fixed. The Martingale Central Limit Theorem uses the requirement of the Lindeburg's CTL and further analyzes that the requirement of independence is for the underlying uncorrelated properties [18], and uses the martingale difference array to construct the necessary condition. Also the Central Limit Theorem for mixing processes, which adapt the bracketing approximation based on a moment inequality for sums of strong mixing arrays [12, 13, 20].

2.2.4 Test for homogeneity

There are many ways to test the homogeneity of the variances, each has their own field of application. We will introduce four of them together with their advantages and drawbacks.

Levene's Test

Levene's test is the most commonly adopted and least to be affected by departure from normality.[4] It uses the F distribution and is relatively simple. However, it tends to be more strict in checking homogeneity and has the tendency to incorrectly reject the hypothesis in some situation. It is defined as:

$$H_0 : \sigma_1 = \sigma_2 = \cdots = \sigma_k$$

$$H_1 : \sigma_i \neq \sigma_j \text{ for at least one pair of } (i, j)$$

Test statistics : given a variable Y with sample size N divided into k groups, where Ni is the sample size for ith subgroup. The Levene's test is defined as follows:

$$W = \frac{(N-k)\sum_{i=1}^{k} N_i (\bar{Z}_{i\bullet} - \bar{Z}_{\bullet\bullet})^2}{(k-1)\sum_{i=1}^{k}\sum_{j=1}^{N_i} (Z_{ij} - \bar{Z}_{i\bullet})^2}$$

[4] There are other versions of levene's test for different behavior of the distribution like skewness or heavy-tailed. We use ... version of the test...

There are three choices for defining Z_{ij}, which determined the robustness of the test according to different situation. where

$$Z_{ij} = \left| Y_{ij} - \bar{Y}_{i\bullet} \right|$$

where $\bar{Y}_{i\bullet}$ is the mean of the *i*th subgroup.

$\bar{Z}_{i\bullet}$ is the mean of Z_{ij} for the *i*th group, and $\bar{Z}_{\bullet\bullet}$ is the overall mean of Z_{ij}.

The significance level used throughout this paper is 95%.

Bartlett's test

Bartlett's test is sensitive to departure from normality. When the distributions tested are not normal, the test may simply be testing normality. While the distributions to be compared are normal, this would be the best candidate. It uses the χ^2 distribution and is defined as:

$$H_0 : \sigma_1 = \sigma_2 = \cdots = \sigma_k$$

$H_1 : \sigma_i \neq \sigma_j$ for at least one pair of (i, j)

The test statistics is to test for equality of variance of the k groups tested. The hypothesis is rejected if the variances for at least two groups are unequal.

$$T = \frac{(N-k)\ln s_p^2 - \sum_{i=1}^{k}(N_i - 1)\ln s_i^2}{1 + (1/(3(k-1)))((\sum_{i=1}^{k} 1/(N_i - 1)) - 1/(N-k))}$$

s_i^2 is the variance of the *i*th subgroup with size N_i. N is the total sample size with k groups. s_p^2 is the pooled variance which is defined as follows:

$$s_p^2 = \sum_{i=1}^{k}(N_i - 1)s_i^2/(N-k)$$

The variance are decided to be unequal if:

$$T > \chi^2_{(\alpha,k-1)}$$

where $\chi^2_{(\alpha,k-1)}$ is the upper critical value of the chi-square distribution with *k-1* degree of freedom at significance level α.

Fmax Test

The Fmax test uses the Fmax table. The advantage of this test is that is simple and quick to do the test. However, it is also affected by non-normality. The Fmax has test statistics as the following:

$$F\max = \frac{\sigma^2_{largest}}{\sigma^2_{smallest}}$$

with degree of freedom being (the number of groups, the data size in each group-1). This is a intrinsive ratio of the largest variance variance comparing to the smallest variances of the groups being compared.

Cochran's Test

Cochran's test is computationaly simpler than the Bartlett's test, but it is also affected by the departure from normality. The test uses Cochran's C table with test statistics defined as:

$$C = \frac{\sigma^2_{largest}}{\sum \sigma^2}$$

the degree of freedom is the same as those in Fmax test.

The Fmax and the Cochran's test differs in their denominator, while the Cochran's test captures the individual variance comparing to the overall variance, which has a similar meaning of the requirement of the Linderburg's Central Limit Theorem, the Fmax test emphesize on the difference of the variances between individuals. Both methods, however, are computationally cheaper than the Bartlett's test and Levene's

test for they make use of the data that can be generated while the subgroups were formed.

2.3 Sampling with replacement

The application of CLT will be based on the probe and probe sets. We will apply the CTL twice to turn the measurements at hand into normal distributions representing the performance of each probe. Detail will be given in the next section. We use a single probe as the basis for forming distributions. Before applying the CLT, we have to ensure the selection of the members in a probe will guarantee the formation of the distribution to be correct. The first time we apply the CTL will be on each probe. Images in each probe have their *specifics* constrained by the property of that probe thus can be viewed as being in the same class. To make the members in the probe to distribute independently we use sampling with replacement for the design of this sampling guarantee that each possible sequence of |p| units has equal probability of being selected [25].

2.4 T Test

A t distribution is a distribution that, given n independent measurements x_i, let

$$t \equiv \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

where μ is the population mean, $\bar{x}$ is the sample mean, and s is an estimator of the population standard deviation defined by

$$s^2 \equiv \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 .$$

The t distribution [8] is asymptotically normal. T distribution is defined as the distribution of random variable when we don't know the real σ. The distribution can be used to draw confidence level or test hypothesis. The t test is used to determine if two *normal* distributions are likely to be the same with respect to the variables tested. The statistics is defined as follows:

$$t = \frac{m_1 - m_2}{\sqrt{A * B}}$$

where

$$A = (n_1 + n_2)/n_1 n_2$$

and

$$B = [(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2]/[n_1 + n_2 - 2]$$

m_i, σ_i, and n_i are the mean, sample standard deviation, and number of data in each group respectively [13]. The confidence level used to test hypothesis in this paper is 95%.

Chapter 3. Experiment Model

In this section we will describe in detail, how to use the mathematics introduced in the previous chapter to construct the experiment. We describe how to use some metrics produced by face recognition algorithms, not necessarily have to be those derived from that defined in section 2.1 (although that is what we will use in our experiment), to form some distribution describing the character of each sensor. Some other techniques that may also be used, but is not included in this thesis, will be mentioned but not explained in detail.

Given certain probe p, defined by some *character*, we can calculate some statistics $\theta_p = T_n(x_1,\cdots,x_n)$ for it. If we get the mean μ of T_n of the images in p by the method described in section 2.3,sampling with replacement, by CTL we know μ will be distributed normally. Let the mean of μ be μ_p and variance be σ_p. This distribution describes the algorithms' response to the character of probe p. Suppose probe p_n is taken by the near camera and p_f is taken by the far camera, and $l_n(e_i) = l_f(e_i)$ where $l_n(e_i)$ designate the label for ith element (image) for the near camera and $l_f(e_i)$ designate the label for ith element for the far camera. We can use the t test to decide if the Gaussian distributions from the two probes are significantly different. The comparison result of the two statistics θ_{p_n} and θ_{p_f} yields the potential difference of the cameras *regarding* the particular probe choice, or the *property* of the

probe. We call two probes p_1 and p_2 are *corresponding* if $\forall i, l_{p_1}(e_i) = l_{p_2}(e_i)$. Two probe sets P_1 and P_2 are *corresponding* if $\forall i,j,k : p_j \in P_1, p_k \in P_2, l_{p_j}(e_i) = l_{p_k}(e_i)$. However, if we wish to see the difference of the two cameras in a more general term, we have to use probes with different properties, best with complementing properties. That is, suppose X is the varying property of the corresponding probe sets P_1 and P_2. $\forall x_i, \bigcup x_i = X, \exists p_k \in P, x_i \in specifics_{p_k}$

Let a series of probes be represented as $P = \{p_1, p_2, \ldots, p_{|P|}\}$. In the series, $l_{p_i}(e_j)$ and $l_{p_k}(e_j)$, where $l_{p_k}(e_j)$ means the label for the jth element for the p_k probe, are not necessarily equal, $\forall i,k, i \neq k, \exists j, l_{p_k}(e_j) \neq l_{p_i}(e_j)$. This means there is no presumed relation between any of the two elements p_i and p_k in P, $p_i$and p_k are *independent* of each other. Remember each element p_i in P is represented by a normal distribution. The sum of several *independent* Gaussian distribution will still be Gaussian. Suppose the mean and variance of element p_i are μ_i and σ_i respectively. We can calculate the Gaussian distribution of P by simply adding the weighted element together; $\alpha_1 p_1 + \alpha_2 p_2 + \cdots + \alpha_{|P|} p_{|P|}$ is normally distributed with

$$E(P) = \alpha_1 \mu_1 + \alpha_2 \mu_2 + \cdots + \alpha_{|P|} \mu_{|P|}$$

$$\text{var}(P) = \alpha_1^2 \sigma_1^2 + \alpha_2^2 \sigma_2^2 + \cdots + \alpha_{|P|}^2 \sigma_{|P|}^2$$

where

$$\alpha_1 + \alpha_2 + \cdots + \alpha_{|P|} = 1$$

Here the two-layered distribution ensured the fulfillment of our two goals mentioned previously. The first layer, the distribution formed from individual images of the probe, reduces the variances for putting similar images in a probe. Note that similar here could be defined differently according to different context. The summation of the distributions of probes in a probe set combines the measurements from different testing sets (probes), and demonstrates the sensor's performance as a whole. Note here that the optional choice of complementing properties of the probes in a probe set makes this representation more complete. The distributions of the individual probes can be used for further analysis of how the varying factor in this probe set affect the recognitions.

By summing the elements of probe set P we are able to both reduce the dimension of the data and possibly reduce the variance caused by different properties' response to image quality and get to know the sensor's behavior as a whole. This serves similar function as the linear discriminate method [9, 10]. Linear discriminate methods basically try to project the data in a way that is best for different data set/classes to be distinguished. We also do dimension reduction to reduce the number of Gaussion distribution representing the performance of the camera to perform t test. However, our goal is not trying to discriminate between the behaviors of the two cameras – we do not

know if there is supposed to be a difference, on the contrary, this is what we are trying to find out. Unlike linear discriminate methods, which lie on determining the weight of the different features, we assigned all α_i to the same value, $1/|P|$. Note that the summation of the weights does not necessarily have to be one, we put one here so that the difference between summation and individual probes can still be seen.

Suppose we have two series of probes, P_f and P_n for each camera, and that P_f and P_n are *equally representative*. We can then compare the two cameras by the student's t test to see if there is a difference in the performances of them.

There are other ways of combing the data from different sets. If we treat P as a $|P|$ dimensional "feature" vector (assuming the selection of a probe set is based on complementing property of the probes), we may preserve the response of the camera to the characteristic of each specification of property. If we are to compare the two cameras, by measuring the distance between the "corresponding" Gaussian distributions, by corresponding we mean the distributions from those probes with *characteristics* only different in the camera used, and $l_n(e_i) = l_f(e_i)$, we can then apply some distance metric like the Hamming distance on these individual features to determine the total difference between the performance of the two cameras in how many or what value of the varying variable. This way we may give a more quantitative approach to the difference between cameras. However, we are to investigate if there is a significant difference between the recognition results due to the disparity of the sensor's performance. The quantitative approach is out of the scope of this thesis and will not be discussed further.

Other distance measurements with confidence level included like the Mahalanobis distance can be used for given statistics of the probes to check the disparity of the sensor's performance. However, the Gaussian distributions formed here can simplify the process without giving up statistical validation.

FaceIt

The algorithm we will use to classify the face images is Visionic's FaceIt. This algorithm claims to improve the shortcoming of the PCA algorithms. This algorithm is based on the method called Local Feature Analysis [17]. The PCA algorithm use the global representation of the features, which is not robust to local changes like occlusion [9, 21]. LFA method is built on PCA with local information that remedies the PCA's inadequacy with low dimensionality. According to Visionics [6][5], their product can recognize human face under various lighting, glasses, ...etc. They also claim to handle the pose variation of up to 35 degree in any direction from full frontal.

The evaluation of the claim was conducted by Gross, etc [9]. Although the result of the evaluation shows the robustness regarding different pose variation and illumination still can be improved. Our goal, however, is not to test the performance of these formations, but is to know what degree does these variations deteriorate recognition results regarding to the reduction of image quality. It would be interesting to

[5] Visionics has merged with Indetix Incorporated, the biggest supplier of biometric technology, in June 26, 2002, and bare the name of Indentix.

know if the deterioration of recognition results differs in the different factors on the same degradation of image quality.

Chapter 4. Experiments

The classifier used to classify the images here is FaceIt [14]. Due to the time-consuming nature of our experiment, we did not utilize other algorithms although same processes can be applied. Images were taken concurrently by the near and the far camera for targets under designated environments. Thus for each property, there are two sets of images having exact context taken by each camera. There are 256 people and four possible poses for each person. We categorize the weather (environment) into three types according to the sky condition: clear, partly cloudy, and mostly cloudy. For each weather condition, data are collected at 243, 137,72 time period respectively. The specification of these four parameters defines the properties of a certain probe.

The face images collected are then fed to the face recognition algorithm. Since the correct classification should have the rank be one and wrong classification can produce rank as large as 256, obviously by taking the rank as it is to compute the Gaussian distribution for each probe can be biased by a small number of seriously-wrong classifications. We use a clamped average to compute the Gaussian distribution: ranks over certain number r are all treated as r. This way the distribution is more likely to show on a whole how sensors really perform. The definition of Central Limit Theorem says when the number of means go infinite, the distribution will approach Gaussian. While we are not able to reach this criterion, different probe sizes might show how our experiments do.

To form the Gaussian distribution for each probe, we have to divide the member of the probes into subgroups and decide the mean for each subgroup. The members of a probe are assigned a subgroup according to their position in the probe. That is, image e_i belongs to $\lfloor i/n \rfloor$, where n is the *sampling size*, or the size of the subgroups in a probe. In our setting, the sizes of all probe sets, probes, and subgroups in a particular trial in an experiment are all fixed. By having different trials we get investigate the influences of the sampling sizes in addition to the hypothesis of that experiment. The constraint for the property (of the probe), however, is the same throughout the experiment. We define the syntax $G(x, y, z)$ to be that the Gaussian distribution for each probe set is of size x (that it consists of x probes), and each probe consists of y subgroups where each subgroup is a data set of size z, the sampling size. Images are collected from October 2002 through April 2003.

The Influence of Image Quality on Different Targets

In this experiment we wish to investigate the how sensor's performance can affect recognition results by looking into the degradation of successful classifications on individual targets. Let the probe set of some camera be P, $P = \{p_1, p_2, \ldots, p_{|P|}\}$ where $|P| = n$ is the number of different targets, which in here is 256, in P. The elements in each p_i consist of images from a particular person. No constraint was put on the pose of that person or the time the images were taken. Images are all taken, however, under clear sky. The person in each p_i is exclusive in the probe set. Thus the property for some

probe p_i in P here would be (T',W',E',O') where T'=DT', W'=Clear, E'=$person_i$, and O'=DO'. The Gaussian distribution from p_i designates the algorithm's performance on this person under clear weather. By simply comparing the Gaussian distributions from p_{fi} and p_{ni}, we get to know how the sensor's quality can affect the recognition result on this particular person. The result in Fig 3 shows the percentage of the targets having better classification results in each camera. Each data point in the graph is the average of ten different probes under the same constraint but independent. Further examining the raw data we found that the targets that the algorithms perform better on the far camera are consistent. That is, some targets are more suitable for recognition when the image qualities are less ideal. In the different formations, which are all of data size (total number of images in one probe) 1500, around 65% of the classes possess significantly different empirical distribution of classification results on the two *corresponding* probes. This result initially verified the assumption that the difference in image quality can lead to significant difference in recognition results. While the 7% of the results that the far camera have a better recognition results suggests the possible validity of the previous mention effect that some targets are better recognized in blurred images. Different grouping does not affect the proportion of the result as long as the formation is in reasonable range.

To view the changes in recognition results in a more general term, we use the method described in the previous section. Face images from different persons can be viewed as some basic feature being skewed to different degree and tension. The

difference in the Gaussian distributions for P_f and P_n will show to what degree a less-sharp image will deteriorate the ability of the algorithm to identify blurred features. Fig 2 shows the p-value of some targets with different formation. Each target has a stable p-value throughout different formation, which may indicate some targets tend to be less sensitive to the change in image quality on classification. There is no evident relation between the means and variances of the ranks of these targets. The total data size is still 1500. The central limit theorem starts to have a good approximation at sample size around 30. We did not show the result of P_f and P_n in graphs because *all formation has a p-value of less than 0.0001.*

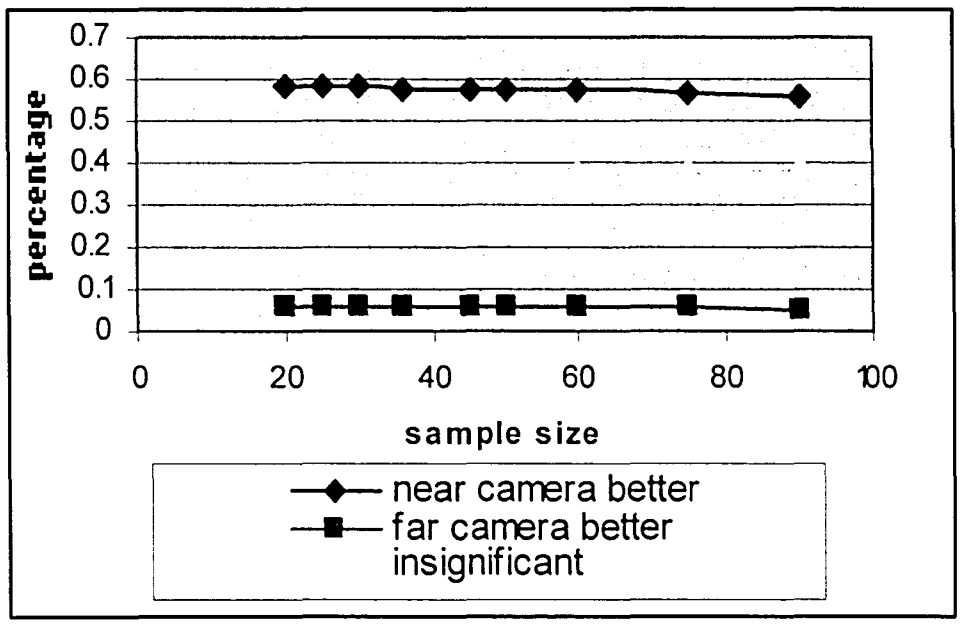

Fig 1. The average percentage of targets' performance relative to different group sizes.

By using different person as the unit for forming distributions, not only do we increase the database by incorporating the different pose and time and past variance test, but also get to further investigate the importance of image quality on individual class defined by different targets. The sum of these individual target distributions results in the empirical distribution of the algorithms' performance in a more general perspective on various targets and, if the database is comprehensive enough, possibly the class of human faces on this camera. All formations of the distributions reject the null hypothesis of the two distributions being identical. From this graph we show that the ability of the sensor can affect the recognition rates to a discernable degree even if the difference is not identifiable by human. Any experiments undergo sensor change should take the effect of the sensor quality into consideration before blindly apply target experiments to ensure the consistency of the results.

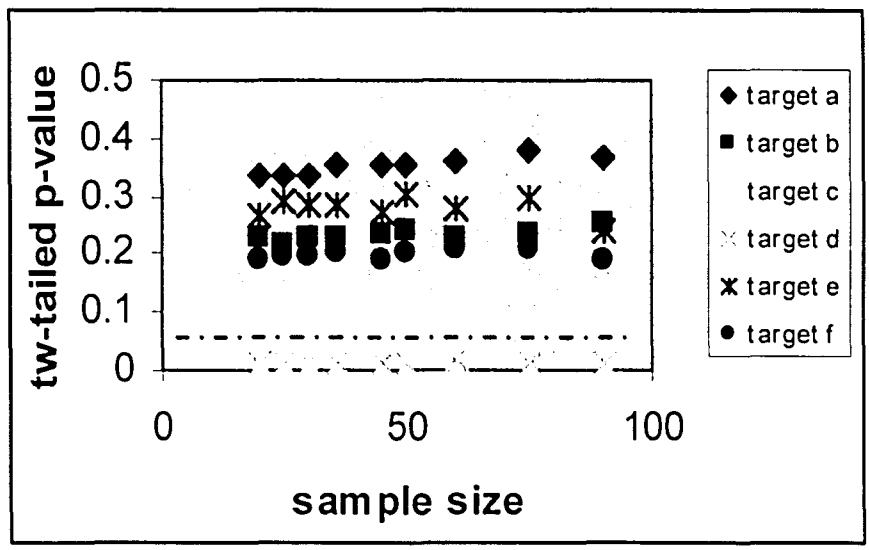

Fig 2. The two-tailed p-value of some targets in different formation of probes.

The Sensitivity of Target's Pose to Image Quality Change Using different poses as the base of individual distributions clarifies the algorithms' ability in recognizing different departure from fully frontal images. Each distribution from a particular pose of the target not only demonstrate the algorithms' ability in recognizing this class of images on that particular camera, but by comparing the corresponding distribution on different camera we also get to know how sensitive this particular gesture react to image quality change. Although this property may be algorithm-specific, this potential discrepancy between classes should be general, with difference only in the quantitative aspect, and similar techniques can be applied to other algorithms as well. This experiment has a similar setting as that in the previous experiment. The difference would be the constrained being the poses instead of target (person). Thus the property for some probe p_i in P here would be (T',W',E',O') where $T'=DT'$, W'=Clear, $E'=DE'$, and $O'=pose_i$. Due to that the data collected only consist of four poses, instead of only having four elements in the probe set P, we can view the definition of probe set as $P=\{p_{1_1},p_{1_2},\ldots,p_{1_n},p_{2_1},p_{2_2},\ldots,p_{2_n},\ldots,p_{4_1},p_{4_2},\ldots,p_{4_n}\}$ where n can be an arbitrarily chosen number so long as it can be used to justify the statistics needed, and that p_{k_i} and p_{k_j}, $i\neq j$, are independent elements of kth pose. This formation is not mandatory and is used to increase the data in each probe set to make it stable?? The different poses impose tilted or missing feature, or features viewed from different angle for the recognition algorithm. The comparison of P_f and P_n then show the algorithm's robustness to blurred version of these already-less-informative features. By summing all

the poses together the distribution is able to represent the performance of the classifier with respect to all possible departure from frontal images on that particular sensor if the database used is general enough.

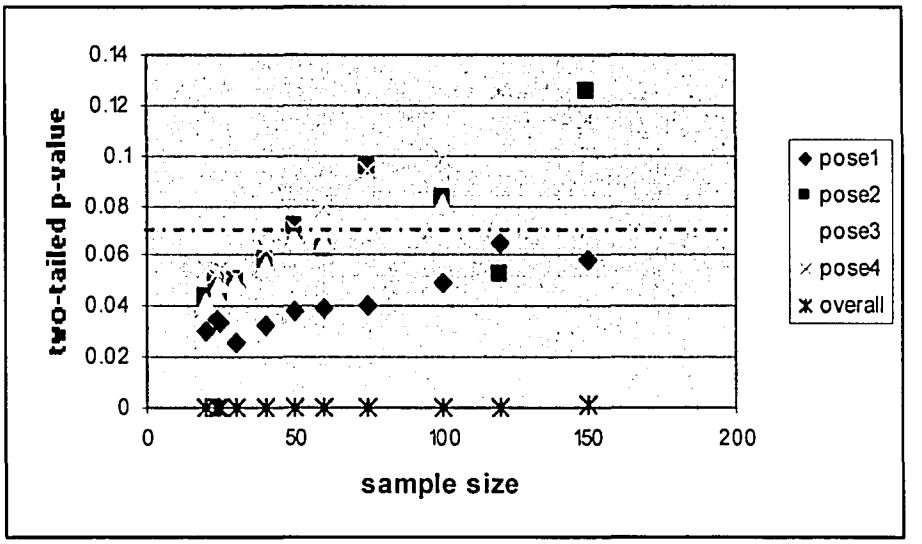

Fig 3. The two-tailed p-value for the distributions of different probes and the probe set.

Fig4 shows the model for each pose. Fig5 shows the two-tailed p-value from the test statistics for distributions of the four different poses and the overall distribution over different formation, where the dash line indicate the 95% confidence level. The increase of p-value actually indicates the decrease of significance level. The p values vary with different poses, which indicated the different degree of impact the image quality has to different poses. While the increase in sample size should improve the approximation of the normal distribution and reduce the variation of the means, the effect that the t value of less degree of freedom has larger p-value overthrow the improvement in

approximation since overall data size has to remain constant. As can be seen some poses tends to be more sensitive to the change in pose, which might be explained by that these poses have more available features for classification and the deterioration of these feature worsen the the overall result, however, the near camera performs better in all poses and are all above the significance level of 99% which indicates a significant impact on the difference of quality on classification results. This consistency may shows that grouping by poses rather than target can eliminate the potential difference

Exploration in Temporal effect

In this experiment the constrained tuple is the time the image is taken. Unlike previous experiments, which deal with the recognition under different treatment of features, that group data different definition of possible classes, grouping by the time the images are taken is dealing with a more general expression of the environment. If the visibility, the atmosphere, even the temperature and all other environmental condition are the same, recognition results of images taken at different time should be distributed around the point decided by other specifics of the images. While the distributions of the pixels of the images should be normal, the pattern of the ranks resulting from this pixel level difference is less obvious. But we can, however, assume the results of the comparison of different sets of data under the same environment defined previously to be quite similar. But this is highly unlikely the case since the environment can't be reproduced

so the context the image is taken can never be identical. The comparisons of the two sensors then tell us how different the two sensors would react to this environmental factor, which might be affected by the different response curve of the cameras. In actual image taking processes, we may record some of the environment fact like the sky condition, wind speed, visibility ...etc. We restrain the condition of the sky to be clear to see how the comparison results can vary under this loose definition of similar environment. The property for some probe p_i in P here would be (T', W', E', O') where $T' = time_i$, W'=Clear, E'=DE', and O'=DO'.

While the p-value of a comparison at a particular time period indicates the performance of the two cameras under that environment, the distribution of the probe set, the summation of the probes, can still demonstrate the overall performance of that camera by smoothing out the temporal effect. Fig 6 shows two set of probes that $\forall a_k \in p_i, b_k \in p_j$, the *specifics* of a_k and b_k $s_a = (t_a, w_a, e_a, o_a)$, $s_b = (t_b, w_b, e_b, o_b)$ have the following relation: $w_a = w_b$, $e_a = e_b$, $o_a = o_b$, and $t_a \neq t_b$ for all p_i and p_j that $i \neq j$. The time constraint for the probes is that it consists of a single element in T, that is, each probe consists of images from a particular time period. We do not call this set of probes as probes sets because it does not fulfill the constraint of a probe set that the probes of a probe set should be independent. The reason for selecting such particular image collection and observe its behavior over time is to make sure the variation of the p-value is not due to the difference in selecting the member of a probe. The results containing the probes of a probe set is also shown in this graph. The graph suggests that during

certain time periods, the p-value does tend to be larger, which makes the difference between the performances of the two probes insignificant. This variation in the p-value under clear sky indicates that although all under clear sky conditions, the environmental variation is still too big to ignore. The probe sets under different formation as in previous section all have a p-value of over 99.9 which indicates the significant performance difference in the two cameras. This result suggests that the data for sensor evaluation should still be collected overtime to smooth out temporal effect and get a more justified result.

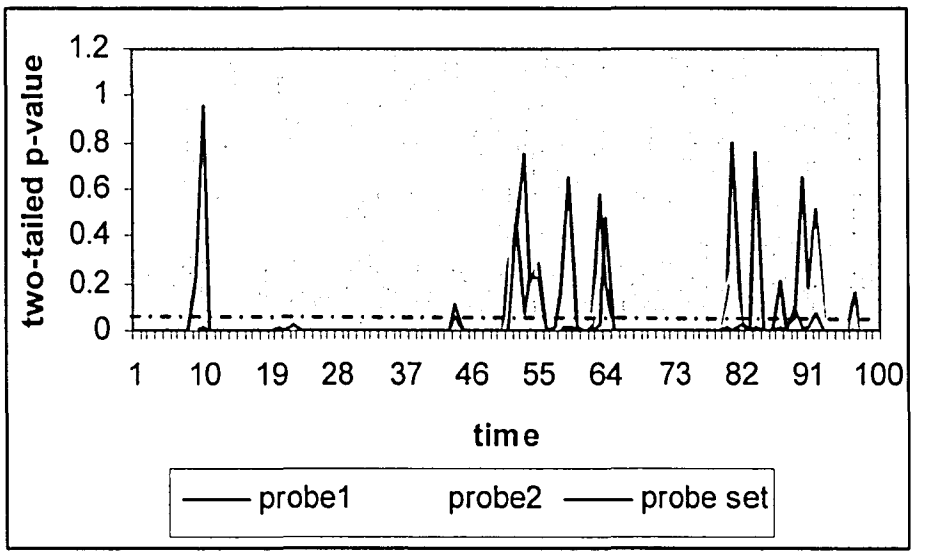

Fig 4. The variation of the two-tailed p-value over time on difference selection of image sets.

Weather effect on the classification

In this experiment we are to investigate if the weather effect can affect the classification to a discernable degree. Unlike previous experiments, which compare the performances of different cameras under similar conditions, we compare *equally representative* probe sets from the same camera under different weather conditions. As suggested in the previous experiments, data should be collect overtime under similar condition to avoid inconsistent result from temporal effects. We divide the sky overcast conditions roughly into clear, partly cloudy, and mostly cloudy to allow enough data to be collected. The formation of the distributions would use pose as the unit for distributions since the poses appear to be more stable/consistent in the distributions in the probe levels, despite that in the probe set level all formation can discriminate the performances of different cameras in significant level. We conduct the comparison on both cameras.

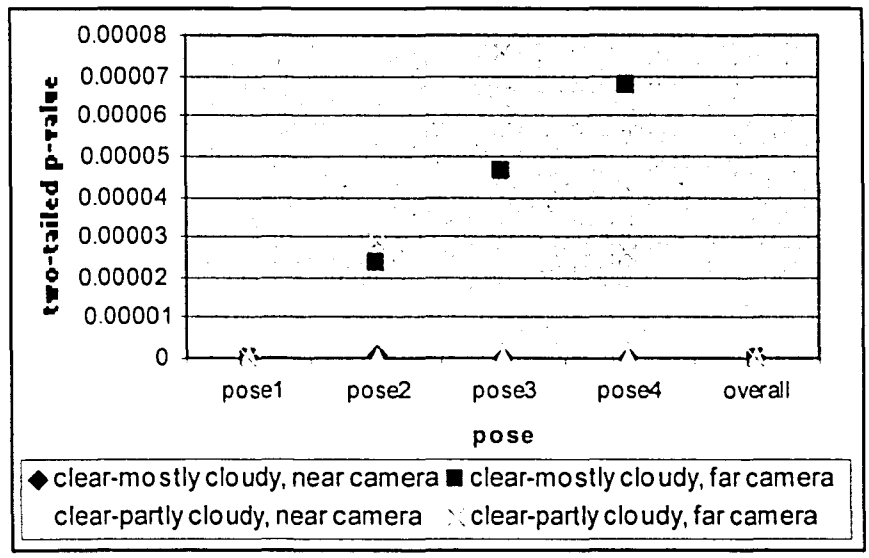

Fig 5. The two-tailed p-value for distributions based on poses under different weather.

While the differences between the distributions from clear sky to overcastted skies from both cameras are all significant, the impact of this weather factor appears to be more apparent on the near camera. Although might be obvious in the figure, the difference of the p-values between sky conditions of both cameras are to the degree of three xxx. The disparity between mostly cloudy sky and partly cloudy sky is insignificant in all poses and overall.

Different definition of Clamped Average

We examine the clamped average using different threshold. Although clamped average should be consistent despite the different formation/threshold, we will investigate to what degree the clamped average will be affect by the settings. Fig 3 shows the clamped average of a threshold of1, 5, and 10. For experiment having threshold being one, the correct classification will receive a score of zero and all others being one. While for threshold being five and ten, the correct classification will have a rank of one, the classification having rank of less than five/ten will keep the original rank, while the results being larger than five/ten will be treated as five/ten.

The result in fig 8 shows consistency throughout the different threshold given. And the tendency of the p-value of different poses also tend to be consistent. That is, in certain pose the p-value tends to be higher despite the threshold used.

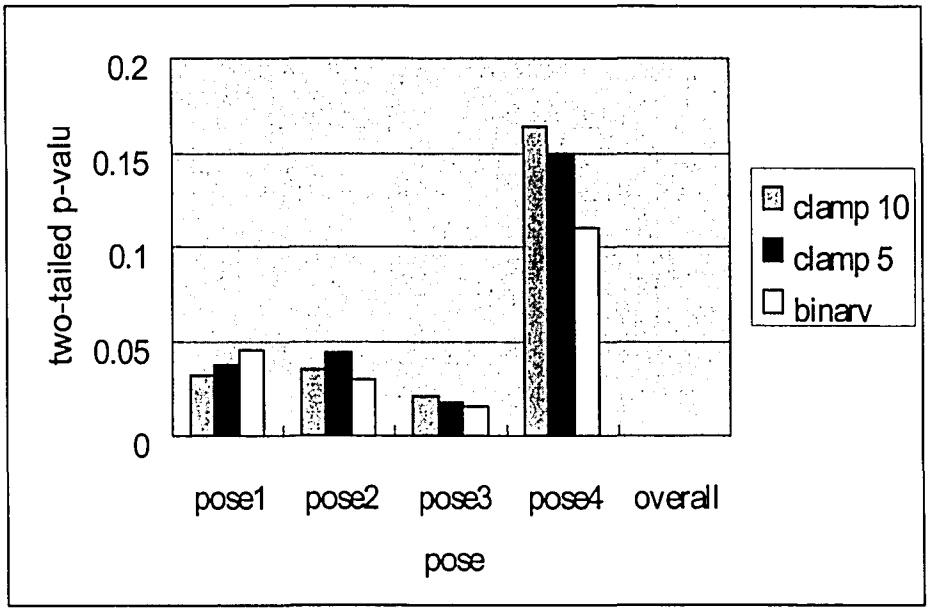

Fig 6. The result of the two-tailed p-value for different thresholds.

Chapter 5. Conclusion

In this paper, we present a methodology for investigating the potential difference of recognition results due to the difference in sensors' performances. This is done by using central limit theorem to force the distributions of the defined measurements of the goodness of recognitions into normal for statistical methods to apply. Further investigation of the parameters that can identify each image in the testing set also reveals how various factors can affect distributions. Despite some inconsistent results in the distributions composed of images having certain specifics in common, the overall distributions composed of a more general set of data shows that there is a significant difference in the recognition results from different sensors. Thus additional care should be taken when sensor change is unavoidable during an experiment to avoid inconsistent results.

References

[1] A. S. Georghiades, P. N. Belhumeur, D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose", IEEE Transaction on Pattern Analysis and Machine Intelligence, Volume: 23, Issue 6, pp 643-660, June 2001.

[2] Baback Moghaddam, Wasiuddin Wahid, Alex Pentland, Beyond Eigenfaces: probabilistic Matching for Face Recognition, Proc. of Int'l Conf. on Automatic Face and Gesture Recognition(FG'98)

[3] Berk, K. N. (1973). A central limit theorem for *m*-dependent random variables with unbounded *m*. *Ann. Probab.* **1**, 352-354.

[4] Binglong Xie, V.Ramesh and Terrance Boult, "Sudden Illumination Change Detection using order consistency", Workshop on Statistical Methods in Video Processing, June 2002.

[5] D. J. Beymer, Face Recognition under Various Poses, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp 756-767, 1994.

[6] D. M. Blackbum, M. Bone, and P. J. Phillips, Facial Recognition vendor

[7] Feller, William, An Introduction to Probability and its applications, 3rd Edition, Wiley, New York.

[8] Fisher, R. A. "Applications of 'Student's' Distribution." *Metron* **5**, 3-17, 1925

[9] Gross, Ralph, Yang, J., Waibel, A., 2000. Face recognition in a meeting room. In: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition. Grenoble, France

[10] Laurenz Wiskott, Jean-Marc Fellous, Norbert Krüger, et al, Face Recognition by Elastic Bunch Graph Matching (1999), Proc.\ 7th Intern.\ Conf.\ on Computer Analysis of Images and Patterns, CAIP'97, Kiel

[11] Lindeburg, J.W. Eine NeueHerleitung des Exponentialgesetzes in der Wahrschein-lichteitsrechnung. Mathematiche Zeitschrift. 15, 211-225

[12] M. I. Gordin, The central limit theorem for stationary processes, Soviet Meth. Dokl. 10 (1969) 1174-1176

[13] M. Rosenblatt, A Central limit theorem and strong mixing conditions, Proc. Nat. Acad. Sci., vol. 4, pp. 43--47, 1956.

[14] Michaels, R.J.; Boult, T.E: Efficient Evaluation of Classification and Recognition Systems, Computer Vision and Pattern Recognition, pp I-50 – I-57 vol. 1, 2001.

[15] Moghaddam B., Nestar C., and Petland A., Bayesian Face Recognition Using Deformable Intensity Surfaces, IEEE Conf. on CVPR, 1996

[16] Oscar Kempthorne, Leroy Folks, Probability, Statistics, and Data Analysis, Iowa State University

[17] Penio S. Penev, Joseph J. Atick local Feature Analysis : A general statistical theory for object representation (1996)

[18] Phillips P.C.B. and Solo, V. Asymptotics for linear Process, annals of Statistics, 20, 971-1001

[19] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss: The FERET Evaluation Methodology for Face-recognizing Algorithms, IEEE Trans. Pattern Recognition and Machine Intelligence, 22(10): 1090-1104, October 2000.

[20] Pollard, D Convergence of Stochastics Processes, New York: Springer

[21] Ralph Gross Jianbo Shi Jeff Cohn, Quo vadis Face Recognition? CMU-RI-TR-01-17 June 2001 Robotics

[22] Richard. O. Duda, Peter E. Hart, David G. Stork, Pattern Classification, 2nd Edition, Wiley-Interscience.

[23] Romano, J.P. and Wolf M., A more general Central Limit Theorem for *m*-dependent random variables with unbounded m, Statistics and Probability Letters 47, 115-124

[24] Sipser, Michael, Introduction to the Theory of Computation, 1st Edition, Brooks Cole.

[25] Steven Thompson, Sampling, 2nd Edition, Wiley

[26] T. Shakunaga, K. Shigenari, "Decomposed Eigenface for Face Recognition under Various Lighting Conditions", Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001

[27] Wenyi Zhao, Arvindh Krishnaswamy, Rama Chellappa, Daniel L. Swets, John Weng, Discriminant Analysis of Principal Components for Face Recognition, 3rd International Conference on Automatic Face and Gesture Recognition (19998)

Vita

The author, Sui-Yu Wang, was born in Taipei, Taiwan, on March 23, 1978. She is the second child of the family with one sister. Her father, Tai-Hui Wang, owns a trading company and runs his business mainly in Mexico and south-America. Her mother, Pei-Shia Chen is a junior high school teacher in the Taipei County. Her sister, Sui-Chia Wang, works in the United Parcel Service as Financial Specialist after she got her MBA degree from the Mendoza College of Business, University of Notre Dame. Ms Wang got her bachelor degree from the department of Computer Science and Information Engineering, National Taiwan University. After her graduation from the university, she worked as an adjunct research assistant in the Institute of Information Science, Academia Sinica. She went to the United States for graduate study after one year. She is currently a PhD candidate of the department of Computer Science and Engineering, Lehigh University.

END OF TITLE