

---

## *Face Value*

---

EXAMINING THE FACTORS THAT INFLUENCE  
SUCCESSFUL FACIAL RECOGNITION TECHNOLOGY  
IMPLEMENTATIONS

*July 2005*



## 1.0 EXECUTIVE SUMMARY

---

The myriad claims and counter-claims about the efficacy of facial recognition (FR) systems bewilders many people, even perhaps those in the biometrics technology community. This paper attempts to bring clarity to the topic, reconciling the contradictions arising from a diversity of opinion on this relatively new and very promising technology.

Much of the confusion has arisen from stark pronouncements made by vendors, media, and end-users surrounding the perceived successes or failures of FR implementations. While there are well-established methods for assessing biometric technologies, the implementation of these best practices is rare, and results are often subject to distortion and misuse.

Also included in this document is an outline of the differences between human and machine recognition, and an example that illustrates how context and human experience can taint or distort the perceptions of both.

The paper concludes with recommendations that some quantitative evaluations of FR technologies should be taken with a grain of salt, then points to more useful methods of interpreting the success of FR implementations.

## 2.0 DEFINING FACIAL RECOGNITION

---

“

*...one must be careful to realize that computerized methods of facial recognition...do not recognize subjects in the same manner as a human brain.*

Facial recognition is a form of biometric identification. According to Duane Blackburn, author of *Face Recognition 101: A Brief Primer*, biometrics are “automated methods of recognizing an individual based on their unique physical or behavioral characteristics.”

A key word in the definition is ‘automated’. We take this to mean that the process of facial recognition involves computerized methods to determine identity, using facial features as essential elements of distinction. While that is clearly true, one must understand that computerized methods of facial recognition, even when they work very well, do not recognize subjects in the same manner as a human brain. Although present and painful in the private sector, the problem is particularly acute in the blending worlds of law enforcement, defense and national security. Although information technologies have over the last twenty to thirty years generated countless repositories of valuable data, technological and political disparities prevented simplified access to unified and consistent views of comprehensive information.

### 2.1 Human Recognition

Facial recognition is a biometric we can all relate to. Humans easily recognize faces. However, no human can recognize a long, lost friend from looking at the subject's fingerprint. Nor can they stare at someone's iris, and either verify or identify who they are. Both of these biometrics require automated or manual intervention to interpret the data being presented.

FR on the other hand, is something we all do. It is essentially the only biometric with a human sensory equivalent. For this reason, many people expect computerized FR to work much like the FR in our heads. This is a potentially fatal assumption, and lies at the heart of why expectations of FR technologies are often irrationally elevated.

The processes underlying human recognition capabilities differ substantially from those underlying computerized recognition. No FR system can replicate the enormous recognition capabilities of the human mind. A large part of the brain is dedicated to recognition. Instinctively, our eyes consume information that triggers the firing of millions of neurons as we attempt to characterize entities as friend or foe, edible or inedible. The largest supercomputer on earth cannot replicate this power, and even optimistic technologists doubt such a system could ever be built.

## 2.2 Machine Recognition

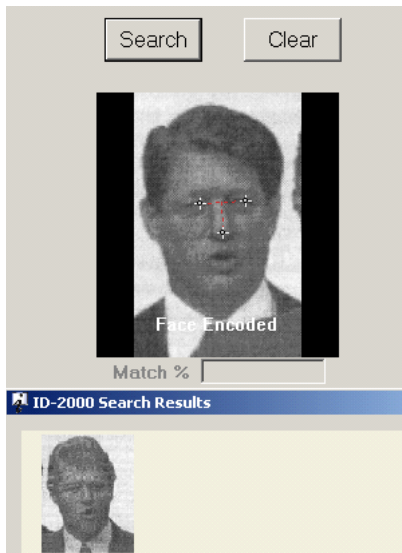
In varying degrees, FR systems deploy some form of mathematical interpretation of a facial image to characterize individuals. Often referred to as coefficients or encode strings, these numbers provide the foundation for automated comparison. Sophisticated mathematics can insert some artificial intelligence or enhanced awareness into the data, but these powers pale significantly in comparison with human abilities to coalesce a lifetime of observation and experience into a single, rapid recognition event.



That said, it should be noted that in reasonably constrained situations, FR technology can actually be superior to human recognition. Consider the image at left.

The majority of persons who view this picture identify the subjects as former US President Bill Clinton and Vice-President Al Gore.

The identification is incorrect. The image is actually two pictures of Bill Clinton, one with a hairstyle and suit modified to look like Al Gore. Most people fail to identify the subjects correctly because humans often use context in a recognition process. In this case, the context distracts from the process because the world is accustomed to seeing pictures of Al Gore in a deferential position and proximity position to Bill Clinton. This leads us to rely on non-facial elements such as hairstyle and clothing to interpret what we see. When this happens, it is easy to make assumptions in the “mind’s eye” about a person’s identity. Context can create a lot of noise in our perceptions, and the brain is easily fooled. This is the guiding principle behind disguise.



People who fail in this little identification test (and almost everyone does) will simply brush it off with a light-hearted smile. It's important to understand that an FR system would not likely make the same mistake, because human assumptions and context have no bearing on the machine's strict mathematical approach to recognition. In this case FR actually does a much better job of determining that the two faces are, in fact, the same individual. The other image at left further illustrates these effects.

The Clinton image, seen in the bottom window, was enrolled into a database of some 6,500 individuals. The database is probed with the Gore image using Visiphor's ID2000 FR. Although almost every single person fails to detect that these are pictures of the same person, the FR algorithm always recovered the Clinton image as the most likely match to the 'Gore' image, effectively determining that they are in fact, the same person.

It may be tempting to suggest that this specific recognition result is fully attributable to the FR product. However, it would be disingenuous of us to do so. In fact, a few caveats are in order.

### 2.3 Facing Facts

We need to consider other factors when assessing the ability of an algorithm to produce sensible matches. Image resolution, similarity of subject pose, lighting equivalence, etc. can have serious ramifications for recognition success.

While it is true that FR effectively determined that the two faces in the Clinton example were highly similar, it is also true that the 'Gore' probe and Clinton database images were very similar in their physical properties. The actual digital pictures had virtually the same resolution, compression level, bit depth, color histogram and even pixel dimensions.

Let's look at the images again. The database into which we enrolled the Clinton image is composed of many different types of facial images. Many are standard mug shots, some are faces clipped from video tape, and others are thumbnail pictures scanned from passports. While we can be pleased with the FR result in this test case, some of the correlation between these two



“

*...All biometric technologies perform better with good quality data, and with consistency between the enrolled data and the probe data.*

faces actually occurred because the images themselves have similar properties. For example, the faces are posed and illuminated in the same ways. If the images had different resolutions, with dissimilar poses and varied lighting, we would expect a deterioration in the correlation between the two faces.

### 3.0 Analysis of Contributing Factors

---

A practical conclusion to these observations is that utilizers of facial recognition systems should endeavor to have facial images conform to a stable and reproducible standard—640x480 uncompressed JPEG as an example. This is why a number of agencies including NIST, ICAO, Immigration Canada, the FBI and the RCMP have recommended generally similar facial imaging standards.

But in our imperfect world, it is quite common and reasonable to base FR projects on large sets of disparate imagery, such as scanned passport photos, ID card thumbnails and so forth. Despite these varying contexts, users still expect a high rate of match success with probe images. As these expectations are extremely difficult to live up to, projects are sometimes labelled as failures despite producing advantageous results and obvious benefits.

An equally common expectation is that surveillance footage of atrocious quality is good fodder for FR systems. While image enhancement and manual intervention in the encode process may assist with FR matching, if the data involves low quality or highly disparate imagery, the results will be unpredictable, and modest at best.

### 3.1 Hardware

In the context of FR, hardware means computers, cameras, scanners and lights. Most companies make recommendations about hardware, and it is safe to say that users should acquire computers loaded with the fastest processors, abundant RAM, high bus speeds, and top quality video capture cards. To engage in a demanding FR project involving live, real-time video capture of subjects with immediate query to a database, there is no point in using anything but most powerful computers and peripherals.

“

*...When someone asks, “What is your system’s FRR/FAR?” they’re not taking into account that the numbers are only meaningful in the context of a specific set of data...*

For smaller FR projects using still imagery queried against a smaller dataset, the hardware demands are lower. Often, budgetary pressures drive elimination of some key, high-end hardware elements, which in turn can compromise project success.

When acquiring faces from a live video stream, there is simply no substitute for using excellent video cameras with superior lenses. Such cameras generally have capable backlight compensation and usable white balance features.

Web cams cost much less than video cameras. That, coupled with their modest size, makes them an attractive buy. Features and functions aside, these cameras are generally deficient in their lenses and other optical components. Most have wide-angle lenses that will add distortion to the dimensions of the face. This effect, even slight, will cause the FR application to struggle with matches. Success should not be expected while using inferior equipment.

We should point out that facial distortion can also occur in still images if the aspect ratio (vertical vs. horizontal pixels) is changed. This is a common problem and images should always be saved with aspect ratio preserved.

### 3.2 Data Quality and Consistency

We have discussed some of the effects that disparate imagery characteristics will have on Facial Recognition. All biometric technologies perform better with good quality data, and with consistency between the enrolled data and the probe data. This is true of FR, in addition to iris recognition and fingerprints.

Anyone enrolled in an iris recognition system knows that the image of their iris, taken upon enrollment, met an exacting standard for position and lighting. Furthermore, the compliance of the subject was paramount. When the subject later probes the system, the conditions for recapturing the iris image are virtually the same as the initial enrollment. Also, the subject must be in complete compliance with the system's demands for presentation. This conformity between probe and enroll conditions provides a reasonable constraint necessary to produce a high certainty of identification.

Also consider that an iris or fingerprint are more stable entities than a face. Human faces change a great deal over time and they are easily altered. Furthermore, human faces can be pitched, yawed, rotated and obscured while still being recognized as a face.

### 3.3 False Acceptance and Rejection Rates

Two measurements are often quoted as identifying the capabilities of FR systems specifically and biometric systems in general. These are the False Acceptance Rate (FAR), and the False Rejection Rate (FRR). These rates refer to the number of false negative or false positive matches returned during a biometric evaluation and verification.

Verification is principally concerned with determining: "Are you who you claim to be?" We must keep verification distinct from the other staple of the biometrics known as identification, a process concerned with who you might be.

In the context of an FR project, questions about FFR/FAR are often the first thing one hears. When someone asks "What is your system's FRR/FAR?" they're not taking into account that the numbers are only meaningful in the

“

*...At the very least, we should realize that the business of verification involves trade-offs.*

context of a specific set of data under clearly defined conditions—then using specific tools for enrolling subjects into a database and later probing it. Changes to any of these elements can result in significantly different FRR/FAR values, rendering the numbers useless in generating an understanding of the system’s capabilities.

Secondly, many FR projects—such as finding suspects in a police mugshot database—are not verification tasks, but are instead identification exercises. FRR/FAR has no application in this case. Measures such as CMC (cumulative match characteristics) are more meaningful.

CMC is a percentage rating that describes how often a probe image is matched against a database of enrolled images, with the correct match being displayed as “most likely.” In other words, if a person’s face on the probe image is correctly identified by the FR algorithm against the enroll database, CMC describes the likelihood that the match will appear in the first, second, third position etc., within a ranked gallery of possible matches.

As an example, the CMC rating of the Visiphor FR Server version 10.0—using highly-compliant, good quality face-forward imagery with no subject expression—is 0.9795 for 20 rank positions (using an internal test database). This means the system will accurately match the probe image with an enroll image and retrieve the matching image within the top twenty positions 98 percent of the time.

Compare this against a worst possible case scenario, with completely unconstrained lighting, pose, expression, resolution and eye wear for both the probe and gallery images. In these situations the Visiphor FR Server experiences a drop in accuracy to a CMC rating of 0.9249 at the 20th rank. This is still a respectable value, but it illustrates how unconstrained environments and data can limit FR performance. It also reflects the commonly-known “garbage in, garbage out” rule inherent in database design.

### 3.4 Just How Accurate Is It?

Most people are not really interested in a numerical measure of the probability of falsely accepting or falsely rejecting a subject. What they typically want is a number that indicates the likelihood of a person being correctly verified, period. This can manifest itself with questions like “How accurate is it?”, or “What’s the percentage match?” Unfortunately, these kinds of questions are enormously generalized and can be fraught with perilous assumptions.

FR or any other biometric would more readily accommodate a question such as “Are you who you say you are and how confident are we of that claim?” Even with a question as clear as this, an answer cannot be evaluated without the analysis of a specific dataset.

If one were to engage analytical methods, an FR verification system could deliver, in a specific dataset, something called the probability of verification. However, to calculate this value you must first determine the FRR, and this can only be done by deriving the percentage of times that a false reject occurs across all individuals, whether there are ten or 10,000 subjects involved.

The need for a quantifiable expression of identification verification is often very strong, but we see that deriving such a number requires some time-consuming mathematics. Most end-users have difficulty interpreting, let alone calculating these figures. Any realistic statement of a probability of verification can only be determined by running tests against a known dataset. These calculations generally result in a graphical output of something called the Receiver Operator Characteristics (ROC), which is actually a plot of the probability of verification against the FAR.

At the very least, we should realize that the business of verification involves trade-offs. In general, as the certainty of correctly accepting someone goes up, the chances of incorrectly rejecting someone will go up as well, and vice versa. Sometimes, the only way to get a high enough level of verification of an individual is to tolerate burdensome levels of false rejects. In other words we can achieve a false acceptance rate of nearly zero, but the trade-off will be

the false rejection of a lot of people. Where this trade-off should lie is highly subjective—dependent on the level of security required and the tolerance of the subjects involved.

As algorithms improve, hardware accelerates, facial imaging standards develop and expectations moderate, we will likely see this trade-off lessen, and results improve.

#### 4.0 CONCLUSIONS

---

Without significant calculation and testing, most FR situations will lack plausible and statistically meaningful FRR and FAR assessments. Vendors and users should be wary of persuasive comments such as:

“We can correctly identify a subject 99.9% of the time.”

or

“Our facial recognition algorithm is pose-independent.”

Many vendors also choose to utilize far-flung FRR and FAR values to numerically bolster their technologies with potential customers, but as we have illustrated in this paper, these numbers are subject to wide interpretation and in some cases, misuse.

Most measures of certainty of verification are usually not arrived at correctly, if at all. They usually end up being nothing more than correlations between the probe image encode string and all the other encode strings in a database. Hence, the returned match in a question of identity simply becomes the image in the database that correlates highest to the probe image encode string.

The issue also clouds the essential message that FR technology—properly implemented and with its constraints fully understood—can play an important role in securing public and private facilities alike. It is a useful tool that should only play a role within a larger security context. When implemented in this way, FR systems provide excellent usefulness and return on investment.

## 5.0 LEADERSHIP FROM VISIPHOR

---

With a proven track record with law enforcement agencies in the United States, Canada, and the United Kingdom, Visiphor offers cost-effective solutions to simplify justice system integration. We also supply specialized software solutions for facial recognition, arrest and booking automation, and evidence tracking databases to more than 200 justice, government and security clients throughout the world. Short-term rollouts, excellent stability, accessibility, security and value are just some of the reasons for our success.

For more information on Visiphor and how we can tailor a solution for your specific requirements, please visit our website:

[www.visiphor.com](http://www.visiphor.com)

or [contact](#) us.