ANIMETRICS

WHITEPAPER:

2D to 3D Mapping Technologies as a Solution to Facial ID Systems

THE SCIENCE OF LIVING MEASUREMENT

EXECUTIVE SUMMARY

Facial identification is poised to become a major component of video surveillance and security systems worldwide. It can present itself as a very enticing method for biometric identification. It is unobtrusive and discreet, and the infrastructure for its deployment is pervasive. Cameras are everywhere, and many private companies, and every government agency keeps photo ID records, the basis for facial identification.

Industry reports estimate that the biometric facial identification market will nearly double every year through 2008. This growth, however, is not guaranteed. Recent public failures of facial identification systems at airports and prominent police departments risk undermining the market's potential. The fundamental flaw in current facial identification systems is that they try to correlate an archived, 2D image of a human face with the complex, irregular, three dimensional characteristics of an actual human face. Faces have irregularly shaped features – noses, lips, ears, hair—that change in appearance as the face turns. In turning, faces also reflect light and produce shadows, essentially creating new images. Many of these visual events and artifacts in the target face cannot be compensated for by the manipulation of data in the archived image, simply because the necessary data often does not exist in the 2D identification template.

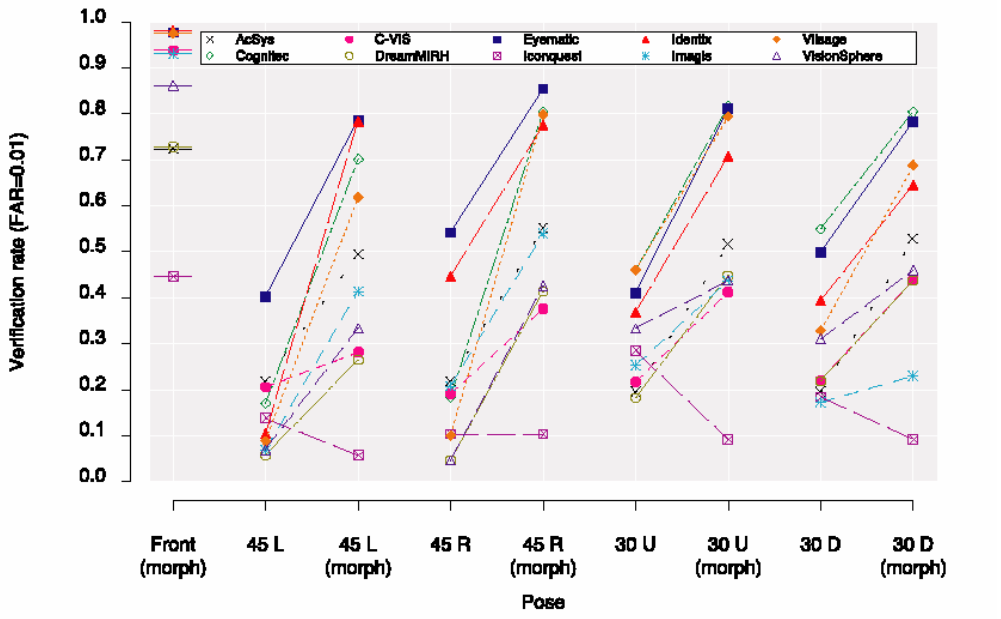

Figure 1: Table taken from FRVT 2002 Report detailing effectiveness of providing corrected frontal image of target image with a face turned 45 degrees to the side, or 30 degrees up and down.

This particular example demonstrates the clear need for pose correction technologies. When conventional facial identification systems were confronted with faces rotated 45 degrees to one side, accuracy rates for identification were consistently under 50%. When a corrected image was provided (see Fig. 2, below), accuracy rates jumped to as high as 90%, clearly demonstrating deficiencies in current facial identification systems. Those corrected images, however, were provided through a manual process, not through an automated system providing in-line correction for identification systems. Additionally, this particular example did not take into account difficult lighting which can be almost as detrimental to facial identification systems.

This white paper explores possible solutions to the limitations of current 2D facial identification systems. We will examine alternative or augmented data capture tools, as well image data processing solutions, including Animetrics 2D to 3D mapping technologies, ultimately demonstrating how Animetrics mapping technologies will be the preferred method for enabling the facial identification market to fulfill its promising growth forecast.

PRINCIPAL PROBLEMS OF 2D FACIAL RECOGNITION

The human face is a complex, irregular object, and the probability of obtaining two identical pictures of the same person, at the same orientation, with the same expression, is virtually zero. When the additional variability introduced by inconsistent lighting environments, camera and lens type, etc., is taken into consideration, the inherent difficulties of precise biometric facial recognition become evident.

HEAD MOTION

The position, or pose, of the head is the first difficulty encountered by facial recognition technologies that rely solely on 2D imagery. Pose can be defined as the head's spatial relationship to the camera. In its simplest embodiment, pose is comprised of six degrees of freedom: three directions of *translation*, or linear movement, (front to back, side to side and up and down); and three modes of *rotation* (pitch, roll and yaw).

Current 2D facial identification systems have no inherent problem with in dealing with the three linear *translations*: front to back, side to side, and up and down. They also deal fairly effectively with the mode of *rotation* known as yaw, in which the head leans from one side to the other. In all of these circumstances the facial features recorded in the archived image, and their spatial relationships with one another, do not change significantly in the target image. Lateral, or in-plane motion, does not change the distance from one corner of the mouth to the

other, or from the left eye to the right eye, all that changes is the positioning of the face and its features in the *image plane*. Likewise, distance from the camera will change the size of the facial features but not their proportional relationship to one another, and it is a relatively easy process to "normalize" an image of a face, or to make the general scale of all faces captured by the camera essentially the same\.

Figure 2 Example of implementation of pose correction system (taken from FRVT 2002 Report), including correction for variations in pitch and yaw.

When head motion causes the face to substantially change its presentation in the image plane, as it does in the two modes of rotation known as pitch and roll, variations occur in the subject image that cannot be compensated for by the process of normalizing an archived, 2D image. The motions involved in roll (where the head leans forward and backward) and pitch (where the head turns side to side) can literally introduce new information or remove previously visible details. Due to this fact, current facial identification systems, which contain no structural information about the face or head outside of what's visible in the template image, are fatally limited. When a head turns to the side, for example (even as little as 15 degrees), the accuracy of current facial identification systems, which cannot effectively analyze a face which is in a different pose than the template, falls off dramatically. Two-dimensional systems are unable to accommodate new information, such as the previously invisible side of a nose or chin, or to compensate for missing information, such as an invisible or partially occluded eye.

LIGHTING ENVIRONMENT

The other principal condition that confounds current systems is the lighting environment. Even if the target image is in a near perfect position for correlation with the archived image, significant variations in lighting and the introduction of shadows can make identification impossible. The best example of a controlled lighting environment is a television studio, in which variation (from windows, etc.) is eliminated and high intensity lights flood the face, so that all features are uniformly illuminated and all shadows are suppressed. By contrast, facial identification deployments are often — of necessity — in inadequately or inconsistently lit areas: hallways with overhead lights and/or windows introducing variation; outdoor building entry points with weather-related lighting variation; or airport security checkpoints, which may experience any of the variations already described. Overhead lighting creates shadows below the eyes and nose, distinctly changing the appearance of the face. Changes in external lighting can also significantly affect facial appearance: the angle of the sun can cause distorting shadows, and lack of sufficient light can cause loss of detail. Such factors also affect interior sites with numerous or awkwardly positioned windows.

Changes in the lighting environment cannot be estimated or compensated for without an understanding of the *structure* of the face in all three dimensions. Because of irregular shapes in the face it is impossible for current systems to calculate from a 2D image how shadows will be cast by the nose or brow ridge, for example, or how light will be reflected off the forehead. These are critical limitations, and stand in the way of implementing a fully functional facial identification system.

POSSIBLE SOLUTIONS TO 2D FACIAL IDENTIFICATION LIMITATIONS

The limitations of current facial identification systems stem from a lack of structural understanding of the face, and the needed structural data can only be provided by a 3D approach.

It is clear that for the facial identification market to move forward there needs to be a breakthrough in how identification is performed. Conventional 2D image processing will never provide an adequate solution, and the industry as a whole is settings its collective sights on finding 3D solutions to this problem. The question is: "How can this be done without discarding what attracted us to this solution in the first place?"

The next question is: How should 3D be implemented? Currently two fundamentally different solutions are emerging. There are, on the one hand,

systems which utilize alternative or augmented data capture. These include stereo-camera arrays, projected grid systems, structured lighting systems and laser rangefinders. Each of these either supplant or supplement the ubiquitous monocular video camera. There are also, on the other hand, systems which use advanced techniques to extract 3D data from conventional video input, a concept which forms the basis of Animetrics technologies.

REVIEW OF PROPOSED SOLUTIONS:

ALTERNATIVE/AUGMENTED DATA CAPTURE

STRUCTURED LIGHTING SYSTEMS: CAMERA/LIGHT ARRAYS

Structured lighting systems incorporate a conventional camera and an array of lights or flashes surrounding the camera lens. The principal behind their function is that as the lights are projected onto the face from different known angles and in different patterns, the reflections from the face will indicate the shape of the surface. The series of images resulting from the projected light are then compiled into a 3D mesh accurately describing the subject face.

Limitations:
Structured lighting systems require a fairly controlled lighting environment in which to operate. Since they rely on reflections cause by their own lights, they will not operate effectively when there is a significant amount light pollution from other sources.

Another problem is range. Structured lighting systems are most effective at the range for which they are calibrated (and are limited as well by the effective range of their flash array. If the lights cannot be projected onto the subject face with sufficient brightness, or if the subject head is in an unexpected position, the system will be unable to generate a proper model.

Practically speaking, the camera/light arrays and the processors associated with them are expensive—ranging from several thousand to tens of thousands of dollars per station. Although it would be possible to reduce these costs through economies of scale, because of the complexity of this process its cost will never approach that of a conventional video camera.

Another, more substantial practical limitation is that structured lighting systems would really only be useful for enrollment purposes, for capturing archival data rather than for tracking a subject image. Structured lighting systems typically take several seconds to capture a characterization to be used for model generation. This time constraint, as well as their need for a relatively controlled environment, makes them ineffective in a target acquisition mode.

Ultimately, without a method for extracting three-dimensional information from more conventional video or photographic input, structured lighting systems will have only a limited impact on the success of facial recognition systems.

PROJECTED GRID SYSTEMS

Similar in theory to structured lighting systems, projected grids attempt to supplement information provided by a conventional camera. In this case, a grid or series of colored lines is projected onto the face. By calculating the deformation of the projected lines, a fairly accurate 3D model of the face can be generated.

Projected grid systems, although not necessarily as accurate as structured lighting systems, have advantages in that they take an all-at-once approach to data capture and model generation. Unlike structured lighting systems which require several frames of data to collect the information required for model generation, projected grid systems require only a single snapshot.

Additionally, projected grid systems are somewhat simpler in construction compared to structured light systems, relying on two main components (a light and a camera) as opposed to a multitude of lights in addition to the camera. Therefore their cost, both for purchase and maintenance will be lower. However, they will never stand on equal footing with conventional cameras on a cost basis.

Limitations:
Despite speed and cost advantages over structured light systems, projected grid systems have some of the same limitations. Their range is impaired due to the need to project the light grid onto the face. The further away the target is from the camera, the less dense the grid will become. Even with a focusing/range-finding function, the light grid dissolves at extended ranges.

Projected grid systems are also vulnerable to ambient lighting conditions, in that severe shadows or overly bright ambient light could disrupt the measurement of the projected grid.

Cost also remains a factor. Even if the hardware cost becomes nominal, the maintenance costs will be considerable due to the unusual configuration of such systems as compared to the conventional cameras normally maintained by security system technicians.

While a promising technology, projected grids have problems that will be difficult to overcome if the goal is to implement facial identification in a ubiquitous manner.

3D STEREO CAMERAS

3D stereo cameras are the most conventional alternative acquisition device to the standard video camera. The theory is relatively simple. By arranging two cameras in an array, where the relative field of view is known *a priori*, stereo cameras emulate the way people see. They also emulate our depth perception. By establishing a correlation between each camera, it is possible to interpolate the information presented to generate a 3D view.

Limitations:
Since stereo cameras do not rely on supplemental lighting in the manner of structured lighting and projected grid systems, they do not suffer from ambient light conditions in the same way. However, they do have fundamental limitations with regard to effective range. Since each camera's view relies on a corresponding view from the other camera, they are effective only when the face is in full view of each camera and prior to the point where the views of the cameras cross. Additionally, since they also rely on sufficient difference between information provided by each camera, the distance between the cameras is critical. In an array where the lenses are separated by 12", the effective range will be approximately 2'-4'. If the cameras are separated by 3', the effective range will likely be 5'-8'. As the array grows, however, the stereo camera system becomes more and more fragile, with accurate calibrations becoming more difficult. As the array grows, therefore, the cost also grows considerably, both for procurement and for maintenance.

LASER RANGEFINDERS

Laser rangefinders are an emerging technology that holds great promise. An extension of military LADAR technologies, laser rangefinders work by calculating the time it takes for the laser beam to be reflected back to the sensor. They are not affected by ambient lighting and have an attractive effective range.

Limitations:
The primary limitation to this technology is its inability to leverage existing infrastructure in any way.

Cost is also a factor. Current systems are quite expensive, and while this cost is likely to decrease over time, laser rangefinders will always have particular maintenance requirements well above the cost of maintaining conventional security camera systems.

Laser rangefinders also tend to be relatively delicate, especially when installed in arrays with conventional cameras, where the 3D geometry is matched to the texture data provided by the camera. As with the camera/light arrays of structured lighting systems, or with stereo cameras, making two pieces work together in a precise manner will always be a relatively delicate marriage.

COMMON PROBLEMS NOT ADDRESSED BY AUGMENTED DATA CAPTURE

One issue not addressed by *any* alternative/augmented data capture system is the remaining need to process the new data. Just because a method for acquiring a 3D model has been determined, the method remains powerless without mechanisms to properly manipulate the new information.

A 3D model generated by any of the previous devices is still going to contain shadows and other lighting artifacts, and will therefore require systems to remove these artifacts from the avatar. And in addition to lighting, the problem of pose remains: it will be necessary to be able to calculate robustly the pose of the face, and the solution to this requirement is not apparent in any of the augmented/alternative data capture systems. In spite of their improved approach to data capture, all of these solutions remain dependent on the development of sophisticated software mapping technologies.

A further issue with all of the 3D capture solutions is the lack of any inherent mechanism for updating legacy enrollment data. This will be a critical limitation the goal is the augmentation of current facial identification systems, or leveraging of large databases if identification images. None of these systems will provide a practical solution to either need.

IMAGE DATA PROCESSING SOLUTIONS

We will say it again: one of the most attractive features of facial identification is the ability to rely on a pervasively deployed infrastructure (video surveillance cameras). With that in mind, methods for augmenting the usefulness of ordinary video data are being explored. Rather than changing the data capture device (the camera), as in the alternative/augmented data capture systems, data processing solutions do just that: implement improved methods of processing already existing databases and video cameras.

IMAGE-PLANE LANDMARK SOLUTIONS: DEFORMABLE GRIDS

Using advanced mapping algorithms, a major objective of these technologies is to derive 3D pose data from the 2D image stream. Most systems in this category use a landmark-based approach, identifying relevant points on the face. Most

also use a deformable grid which constrains possible landmark locations. Utilizing the relative position data of each landmark, the pose of the target is obtained.

Limitations:
Since the deformable grid is a flat template, it does not inherently provide depth information about the target. Therefore such systems have a limited range of positions and orientations within which they are effective. Utilizing a flat template means that occlusion will not be recognized effectively. As the head turns away from center position, and one eye or the corner of the mouth becomes partially or completely hidden by the rest of the face, these systems have to "guess" based on a calculated momentum of motion where the face is going. If the target moves in the manner predicted by the system, data collection continues, but if the target moves in an unexpected manner, the system loses track. Additionally, if the only available target image is in profile, they will likely never obtain the appropriate landmarks in the first place, possibly finding the obstructed part of the face in the background of the scene, providing completely erroneous data for analysis and identification.

Deformable grid systems also share an inability to compensate properly for lighting variation. Again, since the tracking template is 2D, none of the depth-related structure of the face is provided. Without this structural information, complex lighting variation cannot be accommodated. For example, if a strong source of light is projected on the right side of the face, a strong shadow will correspondingly be cast over the left side. In this circumstance the 2D template is likely to fail because it cannot rely on the depth-structure of the face to generate an appropriate lighting scenario.

In summary, while image-plane landmarking systems can be effective for adjusting for minor variations in pose (~<25 degrees), they become ineffective for making corrections for a greater angle of rotation. Additionally, the lack of corresponding 3D data to augment the landmarks prevents them from being used as an effective tool for correcting lighting variation.

ANIMETRICS 2D TO 3D MAPPING TECHNOLOGIES: Diffeomorphic Maps via Computational Anatomy

Animetrics mapping technologies do not rely on introducing new information to complement existing video data. Like image-plane mapping solutions, our algorithms leverage existing camera systems to maximize the usefulness of the data provided. However, through an *a priori* understanding of the general shape of the face — an understanding not available to any of the systems outlined above — we are able to extract the necessary 3D data to correct for pose and lighting variation. This 3D understanding of the face is represented in the

avatar. Avatars carry the full 3D representation of the face, as opposed to deformable 2D meshes which carry only image plane coordinates and are constrained by the relative position of the landmarks.

Animetrics mapping is performed in multiple stages. Using an exclusive rigid motion calculation, we are able to precisely and robustly determine the pose of a head. Once the spatial properties of the face are determined, the geometry is determined and a model constructed through the generation of a diffeomorphic map of the facial geometry. Once the geometry is determined facial motion can be accurately tracked through all six degrees of freedom. Additionally, by using a multidimensional lighting approach that does not presuppose a particular number of light sources, or types of light sources, we are able to normalize the lighting environment, discarding confusing artifacts in the target scene. After pose and lighting have been calculated, it becomes possible to move the face to a neutral position for identification, or continue to track the full motion of the face without concern for occlusion or dark shadows. By tracking in a full 3D, as an eye rotates out of view, for example, the rigid motion system enabled by the avatar remains aware of that occlusion, rather than becoming lost in the process of searching for the missing eye in background imagery. This multi-stage approach contributes to exceptional improvements in accuracy of facial identification.

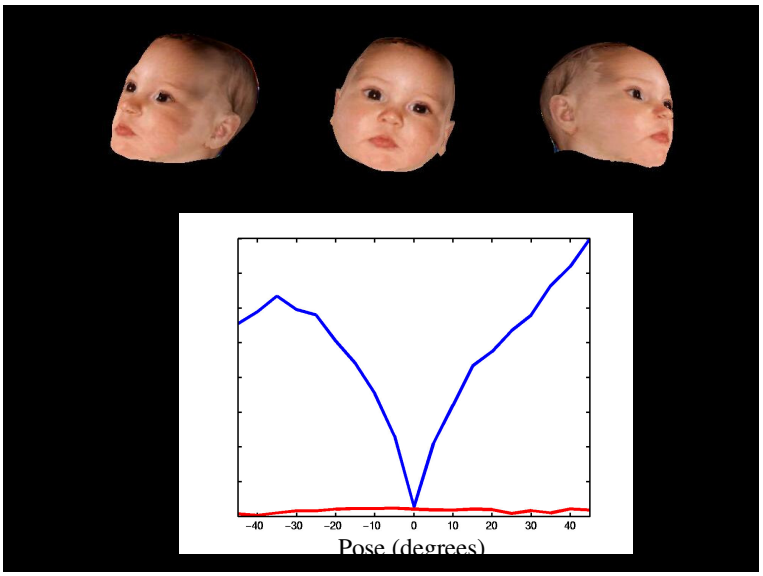

Figure 3 Depiction of squared error calculation comparing frontal image matched to rotated head (blue line) and avatar matched to rotated head (red line) from 0 (neutral) to 45 degrees in each direction

Animetrics mapping technologies have several advantages over the alternative input devices explored above. Rather than requiring the deployment of new infrastructure, Animetrics mapping technologies utilize existing security cameras

without modification or special calibration. Because of our multidimensional approach, our algorithms are able to adjust automatically the particular optical properties of a given camera.

Figure 4 Demonstration of lighting calculated over the surface of the avatar

Another advantage of the Animetrics mapping system lies in presuppositions it is able to make about the input data. Unlike image plane trackers, the Animetrics mapping system begins with the general shape of the face, including depth, and from that it is able to calculate the remaining properties of the face in the target image. Distance from the camera, orientation (or pose), lighting environment--all of these potentially problematic variables are calculated and corrected for, both at registration and during identification. Furthermore, since the success of Since the Animetrics mapping system translates back and forth between 2D and 3D data, these technologies can easily be integrated within the framework of existing facial identification systems and their 2D archival images. Previously collected data is not modified, but rather corrected and optimized.

One issue that is easy to ignore when examining alternative input devices is that no matter how effectively data is acquired, it still must be analyzed. Even if a robust method of data capture is employed, or a robust method of discerning details is implemented in the image plane, the success of image identification remains dependent on technology that can take full advantage of the additional data. Image-plane tracking systems, which provide specific positional details, suffer serious limitations due to a lack of depth structure associated with the provided coordinates. Similar problems occur in alternative input systems: stereo-camera implementations can rather easily acquire a 3D view of the face and produce an image of a face with corrected pose (within limits). However, this system still remains dependent on method for correcting lighting effects. By producing a global solution to these problems, the Animetrics mapping system provides the best possible balance of flexibility in application, leveraging of existing infrastructure (both hardware and software), and power to overcome the limitations of current facial identification systems.

AN EFFECTIVE AND REALISTIC SOLUTION TO PROBLEMS FACED BY CURRENT FACIAL IDENTIFICATION SYSTEMS

Animetrics technologies do not only take advantage of existing ID photographs, and security camera infrastructure. Once our patent-pending technologies have finished their work, the result is still an image. It has been corrected and optimized in ways that 2D technologies could never attempt, but it is the same data that every facial identification system on the market currently uses. Additionally, since analysis and model generation occur completely automatically, there is no need for additional intervention from a technician or an operator. Because of this seamless implementation, our technologies will instantly improve *existing* 2D facial identification systems. Our target acquisition and analysis algorithms simply correct anomalies in the input stream, providing idealized images for more effective facial identification. Likewise, there is no need to alter the registration process, as our analysis and rigid motion detection algorithms will work in-line with existing registration systems. Furthermore, Animetrics technologies are completely backwards compatible with photo-based registration databases, automatically correcting images of previously enrolled individuals en masse. These capabilities translate into accuracy improvements of up to 100% or more[1], with no modification to existing cameras, and with little modification to existing identification software.

Overall, Animetrics technologies will provide the most complete array of image correction and facial identification enhancing systems available anywhere, while also maintaining backwards-compatibility with installed systems.

[1] Based on results published in the Face Recognition Vendor Test 2002 Evaluation Report, March 2003.