Jason Gotlieb
jag2209@columbia.edu
Graduating Senior
May 10, 2012
COMS E6735–Visual Databases
Professor John R. Kender
Final Paper

The Extension of 2D Representations of Human Faces for 3D Analysis and Recognition

Contents:

0 - INTRODUCTION

0.0 *Why Face Recognition Databases?*

Face recognition is an important technology that is becoming more widespread and understood. Databases of faces (or face information) are an important fixture in recognition because they allow computer systems to not only determine the presence of a face but also match, name, and manipulate it under any external condition. In their classic[1] literary survey of face recognition, Zhao et al. attribute two reasons to the noticeable rise in attention towards the study. "The first is the wide range of commercial and law enforcement applications and the second is the availability of feasible technologies after 30 years of research."[2] The first point suggests that face recognition demands a bank of knowledge through which to filter and respond to incoming queries in order to meet the needs of society. In other words, there needs to be visual databases sufficient for computers to perform fully automated tasks. The second point acknowledges that scientists have been working towards this goal since the times of cathode ray tubes. This suggests that there is an intrinsic human fixation with the face. As soon as technology permitted computer imaging, inquiries into face recognition were pursued. Indeed, scientists from multiple disciplines, beyond computer science, have studied human recognition of the face including its structure and representation.

---

[1] In this paper, the term 'classic' will be attributed to academic papers cited by over 2^{10} other academic papers.

[2] Zhao et al., 400.

1- HISTORY AND BACKGROUND

1.0 *A Note on Visual Fixation*

Visual databases have been in the making since before computers existed. The reasons for this are open to debate and scope beyond this paper but are likely due, at least in part, to the fact that humans use a large percent of their brain processing power for vision. Humans tend to spend time seeing things, visualizing things, and processing that which has been seen or visualized. For this reason, it holds that computers should be able to do the same—optimally.

1.1 *An Early Visual Database:*

Sir Francis Galton describes (and then elaborates upon) what could be considered an early visual database in an article published in "Nature" from 1888. The article describes a then existing database that catalogued prisoners by way of measurement classifications. "The primary measures in the classification are four—namely, the head length, head breadth, foot length, and middle-finger length of the left foot and hand respectively. Each of these is classified according as it is large, medium, or small."[3] Sir Galton refers to the French authorities' filing system as a *pigeon hole* of which a prisoner's personal identification card falls into one of the 81 possible *holes*. This system, he explains, "Enables the authorities quickly to assure themselves whether the suspected person is or is not an old malefactor."[4] The French database described envisages future

[3] Galton, Sir F., 174-175.
[4] Ibid., 174.

work of computer scientists who use characteristics of faces to store searchable representations.

Galton's early example is relevant because it tackles the issue of using a three-dimensional query (the suspect) and testing it against a database containing two-dimensional information (the identification card). It is also worth noting that the described French database utilizes an 81-node tree structure to sort the convicts. From today's perspective this prisoner identification system seems crude compared to the many identification and security technologies enjoyed by the developed world. Simple measurements, categorizations, and recognitions however are not an easy task for automated systems. Thus, today's fully automated facial recognition databases resemble the early French model because both models slack compared to modern human perception.

1.2 *Relevant Neuroscientific Background*

The human ability to perceive human faces (among other things) goes unmatched by even the most advanced automated systems. "How neurons encode different percepts is one of the most intriguing questions in neuroscience."[5] While computer-vision-scientists scramble to build an optimal face recognition system, psychologists and neuroscientists use their tools to try and understand the one that comes built in to many human brains. There are times when technologists benefit by modeling nature.[6] It is thus important for computer scientists to keep up with the cutting edge of neuroscience so that a breakthrough can potentially benefit both fields.

---

[5] Quiroga et al., 1106.
[6] Some robots have hands with five fingers for example.

Research by Quiroga et al. exemplifies the approach that neuroscientists take towards understanding face recognition. In their paper—published by "Nature" in 2005—the researchers presented an experimental group with images and words representative of known and unknown people and landmarks. Using chronic depth electrodes, the scientists were able to gather neuronal response to the presented stimuli. They found that, "Cells signal a particular individual or object in an explicit manner, in the sense that the presence of the individual can, in principle, be reliably decoded from a very small number of neurons." So a three-dimensional object found in the world, according to the research by Quiroga et al., is represented in the brain by a subset of cells in the human medial temporal lobe (MTL). If the work by Quiroga et al. is sound, then the brain itself can be understood to operate like a visual database where memory is queried to match real world input.

Galton's early model takes three-dimensional input and queries it against two-dimensional representations—the *pigeon holes* as he describes them contain identification cards, not figurines. The process described by Quiroga et al. involves multi-dimensional input that is queried against representations in the human brain. It has been discussed that the brain encodes representations of this sort in three-dimensions.[7] Three-dimensional representation of three-dimensional objects is the natural structure that visual face databases should thus mimic.

[7] See: Reddy and Kanwisher, 410 & Tsao and Freiwald, 391 and 393.

1.3 *Classic Papers in Object and Face Recognition:*

1.3.1 *Parametric versus View-Based Eigenspace Sets*

Cutting edge technology is built upon the accomplishments of earlier work. Much of the work in object recognition is based on the use of eigenvectors. Turk and Pentland argue for the case eigenfaces in their classic paper from 1991. Refer to Box 1 to understand an early algorithm for creating and maintaining an eigenface database as described by Turk and Pentland. Murase and Nayar discuss the problems faced by computer vision scientists in their 1993 (revised and republished in 1995) paper about face recognition. The paper aims to present a method for, "Automatically learning three-dimensional objects from their appearance in two-dimensional images."[8] This early attempt at object recognition is chiefly relevant for two reasons. The first reason is that it demonstrates the abilities of eigenvectors. And the second reason is that the paper characterizes automatic gathering of image sets. Objects are placed on a turntable that rotates while a camera captures multiple images of them. These two-dimensional images of three-dimensional objects constitute the database and are later used to calculate eigenspace.

At the time of writing, the accepted belief was that, "The human visual system represents objects by a set of two-dimensional views rather than a single object-centered three-dimensional model."[9] Since human observation and modeling is a major informant to computer vision, it follows that Murase and Nayar would have attempted to mimic brain behavior in their work. As information develops however—twenty years later—

[8] Murase and Nayar, 6.
[9] Ibid.

two-dimensional representations are not considered cutting edge in human or computer representation.

It should also be noted that Murase and Nayar use parametric eigenspace, a method that considers shape before angle. Because of this, "Consecutive images are strongly correlated" [10] and, "Their projections in eigenspace are close."[11] This correlation is seldom changed by differing views of the same object. Slightly later research by Pentland et al.—specific to face recognition in a large database—will demonstrate that a view-based eigenspace set outperforms the parametric eigenspace method. It is open to debate about which method is better for overall recognition but the case for a view-based set of eigenspaces is strong.

Pentland et al.'s classic paper from 1994 is an example of a piece of research whose fundamental contributions influence much of the current technology surrounding face recognition. The paper briefly characterizes popular face recognition methods of the time and then famously states that, "For real-world applications, we must be able to reliably discriminate among thousands of individuals."[12] For this reason, Pentland et al. test their system against a database containing 7,562 face images of approximately 3,000 different people. Of these people, 128 were used to sample eigen decompositions. "Recognition and matching was subsequently performed using the first 20 eigenvectors."[13] Users query this system by giving face descriptions that are matched to a database in order to return the first 21 closest matches. Subsequent matches can be paged through in groups of 21 faces at a time. The user can also select a face that then queries

---

[10] Ibid., 11.
[11] Ibid.
[12] Pentland et al., 1.
[13] Ibid.

the database again but this time based on the selected face's eigenvector and not descriptors (tags). After testing the system against 200 randomly selected faces from the face database, it found 95% exact matches.[14]

Pentland et al. describe the differences between view-based and parametric sets of eigenvectors. Interestingly, the parametric method—popularized by Murase and Nayar—is similar to the database described by Sir Francis Galton above. It works by creating eigenspaces from NM images. Thus each parametric eigenspace, "Will encode both identity as well as viewing conditions."[15] This is similar in concept to the 81 *pigeonholes* described by Galton. Each *pigeonhole* is synonymous to a parametric eigenspace. A prisoner's measurement (eigenvector) is queried against the database (universal eigenspace). A key difference and innovation presented by Pentland et al. is that a particular individual, N, can be found in various, M, eigenspaces because each eigenspace is determined by view—a variable measure. They favor a view-based method because it can "Yield a more accurate representation of the underlying geometry."[16] Indeed, the view based technique yields to superior performance depending on degree of tilt compared between the views of a given image.

Pentland et al. also extend their view-based method to incorporate what they call *distance-from-feature-space* (DFFS) which is calculated for each eigenvector. DFFS calculations extend the eigenvectors calculated by *sum-of-squared-difference* (SSD); Murase and Nayar used SSD in their parametric eigenspace detection set. Since calculating DFFS is an extension of SSD, each eigenfeature can be effectively computed

[14] Kender 2-16-2012.
[15] Pentland et al., 2.
[16] Ibid., 3.

to detect facial features. This extension however, comes at the cost of computation and each eigenvector that is extended with DFFS calculations require, "one additional convolution over SSD."[17] The crux of the paper by Pentrlan et al. is characterized in Figure 1 that shows the ROC[18] curves of SSD as compared to different levels of DFFS calculations. Deeper DFFS maps imply higher detection and lower false alarm rates. Pentland et al. found that, "The peak performance of the DFFS detector using the first 10 eigenvectors corresponds to a 94% detection rate at a false alarm rate of 6%."[19] As Figure 1 shows, a DFFS with only one eigenvector outperforms base SSD correlations.[20]

In their classic paper, Pentland et al. were able to use a database consisting of two-dimensional representations of information to effectively detect faces with a range of expressions and views. They tested multiple sets of eigenvectors against a large database. Their results were very good but not perfect. A difficulty Pentland et al. met was, "The issue of feature occlusion and feature/background interaction."[21] The former mentioned issue arises because two-dimensional images cannot give the database a full view of the face. The later issue is based in background noise. To solve these problems, Pentland et al.'s system chooses visible features of importance and determine matches from the eigenvectors therein. A three-dimensional representation however would solve the problem of occlusion and provide depth measurements by which to delete background noise.

---

17 Ibid., 5.
18 Heeger 2003-2007.
19 Pentland et al., 5.
20 Information for this paragraph was taken from Murase and Nayar, 12 & Pentland et al., 4-6.
21 Pentland et al., 4.

1.3.2 *Fisherfaces*

In a later, classic paper by Belhumeur et al. in 1997, illumination and facial expression were accounted for but not pose. The authors of this later, classic paper attempt to, "Outline a new approach for face recognition—one that is insensitive to large variations in lighting and facial expressions."[22] The paper proposes a third type of facial recognition, one dubbed, *Fisherfaces*. They test this approach against various eigenface systems and prove that their system outperforms eigen based alternatives. Yet, Belhumeur et al. do not account for pose. Instead, the authors recommend that the reader use the likes of Pentland et al. to recognize pose variation. Again it is clear that three-dimensional representation of human faces is important to face the challenges posed by both pose and illumination.

1.4 *Early 3D Image Representation:*

1.4.1 *Precursors to 3D*

More robust images can yield more complete results in facial recognition. For this reason, databases need to three-dimensional representations of three-dimensional objects. In their paper from "Nature" in 1990, Poggio and Edelman characterize an early network that recognizes three-dimensional objects in a network containing three-dimensional representations. Their impetus for describing such a system attempts to solve the issues encountered by Pentland et al. Specifically, three-dimensional representations address the problem of "variable illumination" and the, "initially unknown pose of the object relative

[22] Belhumeur et al., 711.

to the viewer."[23] Poggio and Edelman presented their work in the late 20th century and thus were limited by technology. They acknowledge that, "The automatic learning of 3-D models is itself a difficult problem."[24] For this reason, they limit their system to the analysis of simple wireframe objects. They also attempt to relate their findings to the human visual system. The system learns the centers of the wireframes and is thus able to model and then recognize them in any orientation. From today's perspective, Poggio and Edelman seem to have presented an oversimplification of the neural mechanisms behind object recognition. Their network is nevertheless effective and relevant to the development of successful three-dimensional recognition.

Early three-dimensional systems of representations of images—as well as systems today—attempt to address the major challenges of recognition: "The illumination variation problem and the pose variation problem."[25] The other major problem is that, "Face images can be partially occluded."[26] Images captured in uncontrolled environments—such as with video surveillance—have unforeseen characteristics. Lighting can create noisy images and facial pose is indeterminable. Any pose has the cost of occluding some part of the face. Only with a full three-dimensional representation can a face be recognized regardless of the environment and the conditions said environment presents.

In 1996, Belhumeur and Kriegman attempted to use three-dimensional linear illumination subspace to eliminate the problem of illumination variance. The scientists developed a system that creates a convex polyhedral cone that describes a given object

[23] Poggio and Edelman, 263.
[24] Ibid.
[25] Zhao et al.,440.
[26] Ibid.

under all possible illumination conditions.[27] The cone's dimensions correspond to the number of surfaces of a given object as determined by illumination. The illumination of an image can then be projected onto an illumination sphere that is partitioned into a number of regions depending on the object's surfaces. Belhumeur and Kriegman showed that a spherical object has a high number of surfaces and therefore its illumination sphere is almost totally banded.[28] This early paper by Belhumeur and Krigeman aimed to answer the question of illumination variance but did not account for pose variations. Still, the illumination cone and accompanying sphere is an important precept to three-dimensional facial modeling.

1.4.2 *3D Face Representation*

Georghiades et al. acknowledge illumination variability and relate that in handling such variability, "An object recognition system must employ a representation that is either invariant to, or models this variability."[29] The authors go on to explain that their proposed method is an implementation of the illumination cone suggested by Belhumeur and Kriegman. They continue, however, by stating that their work is also, "An extension in that it models cast shadows."[30] The cast shadow modeling is important because it is an early example of three-dimensional modeling. The cast shadow model is computed using, "ray-tracing techniques to determine which points lie in a cast shadow."[31] It does this by performing a generalized bas-relief (GBR) transform on the two-dimensional image

[27] Belhumeur and Kriegman, 245.
[28] Ibid., 256.
[29] Georghiades et al., 52.
[30] Ibid.
[31] Ibid., 55.

whereby an extra plane (third-dimension) is added with the effect of flattening or extruding characteristics. Figure 2 shows an example of a cast shadow model as described by Georghiades et al. The paper thus offers a more complete solution to the problem of illumination variance through the use of three-dimensional modeling. By using illumination cones and modeling cast shadows, the system presented by Georghiades et al. outperforms all other models reliant on two-dimensional representations alone. Figure 3 illustrates how their system reduces error rate in recognizing faces in different illumination conditions compared to other models—including the eigenface model.

1.5 *Overview, Chapter 1*

Facial recognition is important for finding and keeping track of people—and thus requires a large database to query. This chapter took a look at an early database characterized by Sir Francis Galton and demonstrated that even this old and technologically simple system is difficult to automate. It also showed that Galton's early model still has relevant features including its structure.

Faces are also of key interest to the human species. Humans are social creatures that rely on familiarity to interact. Most humans know not to interact with non-human objects or beings because they are hardwired to differentiate between things. There is a trend in computer science to mimic human activity and so the implication that there is a three-dimensional representation of objects in the human brain calls for similar computer representations.

This chapter looked at three classic papers that relate to facial recognition. Each paper tackled the problem of recognition differently and ingeniously. All papers found shortcomings in their systems that were based either in image illumination or pose variances. Solving or coming close to solving one issue was seemingly at the expense of another. Still, it is clear from this overview that technology is rapidly approaching a threshold through which facial recognition will be able to tackle both problems at once. As scientists reach this threshold, there is a trend to represent images with three-dimensions.

The final section of this chapter looked at an early three-dimensional facial representation system developed by Georghiades et al. Their system relied on an illumination cone to create a cast shadow model that greatly improved facial recognition under varying angles of illumination. The early three-dimensional system did not account well for pose variances but still outperformed all previously discussed models.

This chapter contains three accompanying figures and one information-box. Readers of this chapter will clearly see that facial recognition depends heavily on databases and the way information is represented within those databases. Readers will also understand that in order to prepare a facial recognition system to deal with uncontrollable input conditions from the natural world, highly correlated, three-dimensional representations of faces must be analyzed and stored in the database.

2 – 3D REPRESENTATIONS OF 3D FACES

2.0 *Answering Questions with Questions*

With the progression of technology, face recognition too progresses. Yet, state of the art technology—specifically, state of the art face recognition technology—still has vast opportunity for development. Three-dimensional[32] representations of human faces offer an improved solution to the major issues of facial recognition—they are: facial pose variance, facial image illumination variance, and internal facial expression variance. In their chapter, *3D Face Recognition*, Gökberk et al. explain that 3D face information "Is inherently insensitive to illumination and pose changes."[33] The, albeit simplified, reason for this is that the problems of illumination and pose have to do with background (light) and edges respectively. 3D representations delete background, and model 360-degree views of the head—thereby deleting the illumination problem and factoring for every possible head pose. Interestingly, and to technologists' dismay, facial expression is not easily solved with 3D technology.

With progress, a new set of issues arise that must be acknowledged. Gökberk et al. explain that, "The drawbacks of 3D face recognition compared to 2D face recognition include high cost and decreased ease-of-use for sensors, low accuracy for other acquisition types, and the lack of sufficiently powerful algorithms."[34] In other words, getting 3D face information into a database is both expensive (high monetary cost and high computing cost) and complex or else it is hopeless. Beyond this, the algorithms of today do not perform as well as one might hope. 3D face recognition does have the

---

[32] In Chapter 2, dimensionality will be referred to with a variable, N, representative of a given dimension (e.g. 2D, 3D, ND).
[33] Gökberk et al., 263.
[34] Ibid., 264.

potential overcome feature localization but it is not an inherent feature. Some face recognition systems combine the state of the art benefits of 3D with the breadth of research in 2D face recognition systems—they are called multimodal systems.

2.1 *Multimodal Face Recognition Systems*

In their 2006 paper, Bowyer et al. compare multimodal face recognition system with multisample systems. They define their terms by explaining that they, "Compare the performance improvement obtained by combining three-dimensional or infrared with normal intensity images (a multimodal approach) to the performance improvement obtained by using multiple intensity images (a multisample approach)."[35] In sum, Bowyer et al. found that among multimodal samplings, a combination of 2D and 3D systems performed the best. The best multisample system outperformed the best multimodal system though. The winning multisample condition, however, was based on a specific facial expression and lighting condition.[36] For this reason, it is clear that the winning multisample system was the best because it sampled the data with the least noise. In short, the multimodal data that combined 2D and 3D systems as presented by Bowyer et al. outperforms all other tested systems except that system that uses the best possible face expression and lamination. For this reason, the data seems inconclusive and unsound.

---

[35] Bowyer et al., 2000.
[36] Ibid., 2008.

2.2 *Using SVMs:*

2.2.1 *For 3D Facial Recognition*

In early 3D rendering of faces for recognition, achieving pose and illumination invariance was somewhat of a novelty. Huang et al. use component-based recognition and 3D morphable models—technology that was then (2002) considered new technology.[37] "The 3D morphable model is used to generate 3D face models from only two input images from each person in the training database."[38] The 3D face models are used to train the component based support vector machine (SVM) for use in facial recognition. The component-based detector performs, "The detection of the face in a given input image and the extraction of the facial components which are later needed to recognize the face."[39] It did this by generating approximately 7,700 synthetic faces extracted from the 3D head models (hence *morphable*). The synthetic faces each had varying poses and illuminations. The synthetics were then used to train the facial detection and recognition system by identifying different components of interest within the model. The morphable 3D head model informed the system of interesting coordinates in the input images. These coordinates were used in extracting the components—such as eyes, nose, and mouth. All the components are then used to inform the automated face detection and recognition. The resulting component-based system is compared to a whole face system. The ROC curve in Figure 6 clearly shows the superiority of this component-based system as presented by Huang et al.

---

[37] Huang et al., 334.
[38] Ibid.
[39] Ibid., 335.

2.2.2 ***For 3D Image Correction***

One issue the facial recognition system proposed by Huang et al. had was that the 3D head models had holes in them. That is to say that because of noise in the input images (possibly caused by poor equipment or human error), and a generally inferior algorithm (unable to interpolate holes with information), a perfect 3D head representation was not achievable. Technology tends to move quickly and thus in a paper published three years after Huang et al.'s (2005), Steinke et al. propose an system for facial recognition that boasts, "automatic removal of outliers and hole filling."[40] Steinke et al.'s system uses an SVM to convert 3D surfaces so that they are treated like a hyperplane in an ND Hilbert space—where N is a high number and possibly infinite.[41]

The implicit 3D surface proposed by Steinke et al. can then be used for, "smoothing, de-noising and hole filling."[42] It is also extended to create a warp-field that allows it to analyze two objects and, "transform one into the other."[43] Steinke et al. explain that the warp field is important to face recognition because it can, among other things, superimpose or delete glasses. Figure 5 clearly demonstrates Steinke et al.'s system's effectiveness at automatically filling in holes and presenting an overall noiseless and clean 3D facial representation.

Figure 5 is a clear improvement compared to Figure 2 which shows a noisy and incomplete 3D representation. It is worth noting that the ease with which a reader might differentiate between Figures 2 and 5 is testament to the immense capabilities of the human visual system. Figure 2 was generated in 1998—it took technology 7 years to

40 Steinke et al., 285.
41 Ibid. 285-286.
42 Ibid., 286.
43 Ibid.

generate the likes of Figure 5. Steinke et al.'s system is not tested as a means for facial recognition but its ability to perform warp-field effects between any two objects seems promising for facial detection and recognition. The major downside to their system is efficiency—the 3D faces it generates are expensive to compute.

2.3 *Handling Various Facial Expression:*

2.3.1 *With Deformation Modeling*

Deformation modeling can be extended beyond superimposing sunglasses onto a 3D face representation as was shown in Steinke et al. Lu and Jain use deformation modeling in their 2008 paper that attempts to recognize faces with varying expression. They accomplish this by matching multiview, "facial scans (range images) to 3D neutral face models."[44] Expression variance is thus approached. The deformation model used by Lu and Jain is a generic, fully 3D representation of a face. Even though they acknowledge that, "User-specific deformable models are more robust than the generic deformable model," not every face queried by their system has a designated deformation model. [45] They explain that doing so would be impractical and difficult. Instead, the generic deformation model is used.

In many ways, the generic deformable model acts like a 3D eigenface. It is used to create synthesized 3D models with different expressions. It does this by learning deformations (expression variances) from the deformable model control group which is a subset of the test group in total. Facial coordinates of importance in the deformation model are gathered much like the ones described by Huang et al. They are then projected

---

[44] Lu and Jain, 1346.
[45] Ibid.

back onto the query to test for a match by way of a displacement vector. The effect is similar to the eigenface effect because facial expression is varied computationally. Lu and Jain's generic deformation face model is used to deform the querying image—into multiple synthesized images—and thereby test it against the images in the database. The system by Lu and Jain performs fairly well but based on the ROC curves found in their figures, does not yield overly exciting results.

2.3.2 *With Isometric Model*

A paper from 2005 by Bronstein et al. uses an isometric model that, "Reduces the problem of comparing faces in the presence of facial expressions to the problem of isometric surface matching."[46] The paper claims that the idea of a neutral face, much like the morphing model characterized by Lu and Jain, does not actually exist in the absolute and thus should not be used. Instead, Bronstein et al. suggests using canonical forms that correspond to isometries of a deformable surface (a face for example).[47] The issue with using canonical forms is what the paper refers to as the *embedding error*. The basic question herein is, "Whether an isometric embedding of a given surface into a given space exists."[48] The fear is a decomposition of distance ratios between points of interest. The researchers go on to explain that for complex surfaces, isometric embedding is rarely possible. So, in actuality, the Bronstein et al. use *near-isometric* embedding. Thus the canonical forms, "only approximate the intrinsic geometry of the original surface."

---

[46] Bronstein et al., 25.
[47] Ibid., 12.
[48] Ibid.

Usually fitting a square peg into a round hole doesn't solve a problem but in the case of Bronstein et al., the use of imperfect isometric structures suffices.

Indeed, the authors—two of whom are identical twins themselves—were able to use their system to differentiate between identical twins. They claim that such an experiment is, "One of the most challenging tests for a face recognition algorithm."[49] Their system is impressive because it is able to perform successfully (not optimally) while retaining efficiency. The key to the efficiency is the near isometric embedding. The task of comparing faces is, "Turned into a much simpler problem of rigid surface matching."[50] Bronstein et al., explain that this feature is a tradeoff of speed and simplicity at the cost of accuracy. Still, their system is surprisingly well tuned suggesting that the paper has an overly modest tone. Figure 6 shows the ROC curve that compares the performance of the system developed by Bronstein et al. compared to both the slower, rigid 3D recognizer and 2D eigenface recognition. Their isometric system outperforms both.

2.4 *Cutting Edge*

The state of the art in facial recognition involves 3D image analysis of 3D image sequences. In other words, 3D technology is used to evaluate and recognize movement in 3D video sequences. In a forthcoming[51] paper by Sandbach et al., present a system that does exactly that. The benefits of such work include the existing ability to handle

[49] Ibid., 25.
[50] Ibid., 8.
[51] It's so state of the art that it's not even officially published yet! Note also that— at the request of Sandbach et al.—the citation for this article is slightly different from the format otherwise found in the sources section.

illumination and pose variance. Beyond this, a system that takes 3D image sequences (video) into account can handle, "out-of-plane movement."[52] So faces (and other objects) that exit the frame and reappear with different likeness will not confuse the system. The system proposed by Sandbach et al. uses 3D free-form deformations (similar to the method seen in Steinke et al.) to perform feature extraction. Facial expression is then modeled in full rather than in segments—an improvement over Lu and Jain's system—the proposed method is based on total texture and not action units.[53] These methods are not innovations in themselves but the use of all these cutting edge technologies in one system is novel.

Figure 7 gives an overview of the system in its entirety, note that the last box boasts, "Predicted frame labels."[54] This is an amazing feat considering the total automation and 3D modeling inherent to Sandbach et al.'s system. Sandbach et al.'s 3D system outperformed the 2D test model in almost every respect.[55] Sandbach et al. explain that their system, "Has been demonstrated to exploit the extra information available in the 3D facial geometries to improve on the results found with the 2D."[56] Their state of the art system is an amalgamation of the knowledge in systems that predate it. 3D is an obvious extension of 2D and even though it still has a ways to go, exciting research continues to saturate the literature.

---

[52] Sandbach et al., 1.
[53] Ibid., 2.
[54] Ibid., 4.
[55] Ibid., 8 (Fig. 7 & Table 1), 9 (Table 2).
[56] Ibid., 11.

2.5 *Overview, Chapter 2*

Chapter 2 showed that the knowledge of 2D, as discussed in Chapter 1, is extended and improved with the use of 3D. The reader will feel a sense of excitement and hope with regard to autonomous facial recognition system after reading Chapter 2.

The chapter opened with an acknowledgement of the possibilities and new issues presented by 3D. The major possibilities include the deletion of background noise and lighting to enable the handling of illumination variance. Also, with 3D modeling, pose variance problems are nicely solved. The issue of facial expression variance is a problem that exists across 2D and 3D representations. With innovation—Chapter 2 noted—3D representations of faces pose new problems. These include high computational cost and overall complications on all levels of algorithmic construction and implementation.

Chapter 2 took a look at some specific systems and ideas that attack the problems of facial recognition *head on*.[57] Multimodal systems were shown to be promising but as with anything new, have their own problems. The use of machine learning, specifically through SVMs, was shown to be effective at 3D facial recognition and 3D representations in Chapter 2, section 2.

Chapter 2, section 3 dealt with the issue of variant facial expression. Two different models were proposed that represent a fine sampling of the possibilities of extending 3D representations to recognize various facial expressions. The two proposed models use deformation modeling and an isometric model respectively. The former model recognizes expression as an amalgamation of cued parts. The later looks at the

[57] Pun intended.

face as a total surface where each expression is a new surface that is compared to other whole surfaces.

Chapter 2, section 4 took a look at the cutting edge. 3D is moving towards full 3D video recognition that enables it to determine out-of-frame motion. 2.4 demonstrates that technologies grow out from one another and that combining working principals leads to greater innovation. Chapter 2 contains 4 figures.

3 – CONCLUSION

3.0 *Thoughts and Forecasts*

Computers are extensions of the human brain. They aim to aid and optimize tasks that humans iterate daily. Computers know how to calculate the speed our car is traveling and when to flush the toilet for us. As they progress, they are seamlessly integrated into society thereby becoming less visible. The downfall of computers however is not their invisibility but rather their inability to see. With sight, computers will be able to drive for us or let us know when we try to walk into the opposing gender's bathroom. Facial recognition is the first step towards such functionality. Recognizing a face won't automatically enable cars to drive themselves but at the same time, a car will never drive itself if it doesn't know the difference between a traffic light and a pedestrian. For this to become a reality, comprehensive (large) databases must be utilized and manipulated.

An ideal facial recognition system will combine the ambitions of Sandbach et al. with the beautiful representations of Steinke et al. This ideal system will be able to delete illumination variances and handle all possible pose configurations. In addition to this, the system will be able to model warp-fields between any given facial expression so that recognizable as well as intermediate expressions are handled. The system will rely on a database containing full 3D head representations and 3D movement sequences. Just as Belhumeur and Kriegman showed the possibility of modeling all illumination conditions, the ideal system would model all possible face distortions. With computer vision, and detailed facial representation, lips can be read. This would be a huge progression whereby deaf people could better communicate and interact with the world.

The database required for such a system should model the human neural pathway. For this reason, there must be a stronger collaboration between computer vision scientists visual neuroscientists. The brain and human visual system is flawed but highly suitable. Reverse engineering the likes of this is extremely ambitious. One other idea that might lead to such functionality would be to use a "Mechanical Turk" like system to collect vast amounts of data recorded with binocular cameras. The reason such an idea is attractive is because many of the studies discussed here test their data on relatively small databases. It is important to oversimplify things in gathering basis information but at some point, the power of 3D computer facial recognition needs to be put into the hands of the masses.[58]

[58] Thank you.

4 – BOX AND FIGURES

Box 1
Facial Recognition Using Eigenfaces

Eigenface Initialization Operations:
1- Acquire an initial set of face images (the training set/ visual database).
2- Calculate the eigenfaces from the training set, keeping only the M images that correspond to the highest eigenvalues. These M images define the face space. As new faces are experienced, the eigenfaces can be updated or recalculated.
3- Calculate the corresponding distribution in M-dimensional weight space for each known individual, by projecting their face images onto the "face space."

[These operations can also be performed time to time whenever there is free excess computational capacity.]

Recognizing New Faces Images:
1- Calculate a set of weights based on the input image and the M eigenfaces by projecting the input image onto each of the eigenfaces.
2- Determine if the image is a face at all (whether known or unknown) by checking to see if the image is sufficiently close to "face space."
3- If it is a face, classify the weight pattern as either a known person or as unknown.
4- (Optional) Update the eigenfaces and/or weight patterns.
5- (Optional) If the same unknown face is seen several times, calculate its characteristic weight pattern and incorporate into the known faces.

Turk, M. and Pentland, A. *Eigenfaces for Recognition*. Journal of Cognitive Neuroscience, 1991. 3, 73.

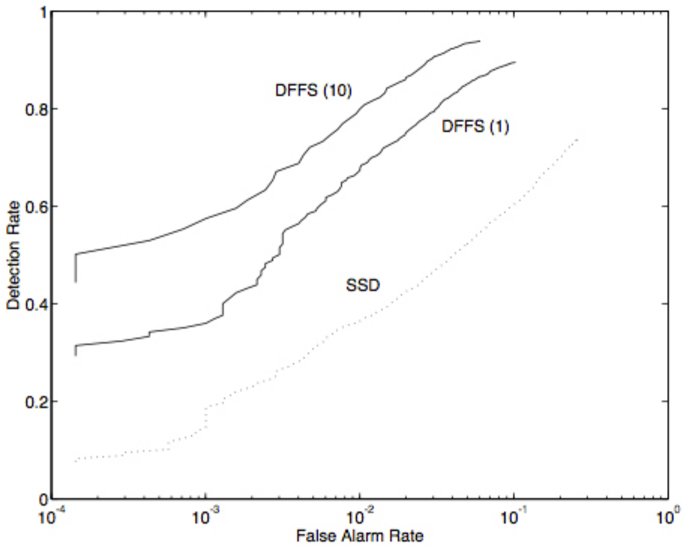

Figure 1:
ROC curve for left eye using DFFS detectors with 1 and 10 eigenvectors. Each DFFS level requires an extra correlation. An SSD detector is shown for comparison—a similar model was used by Murase and Nayar. (Pentland et al., Figure 6, page 6)

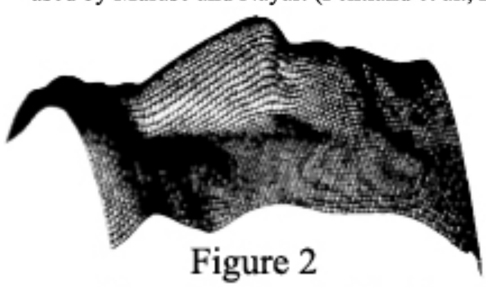

Figure 2
Georghiades et al., 56.

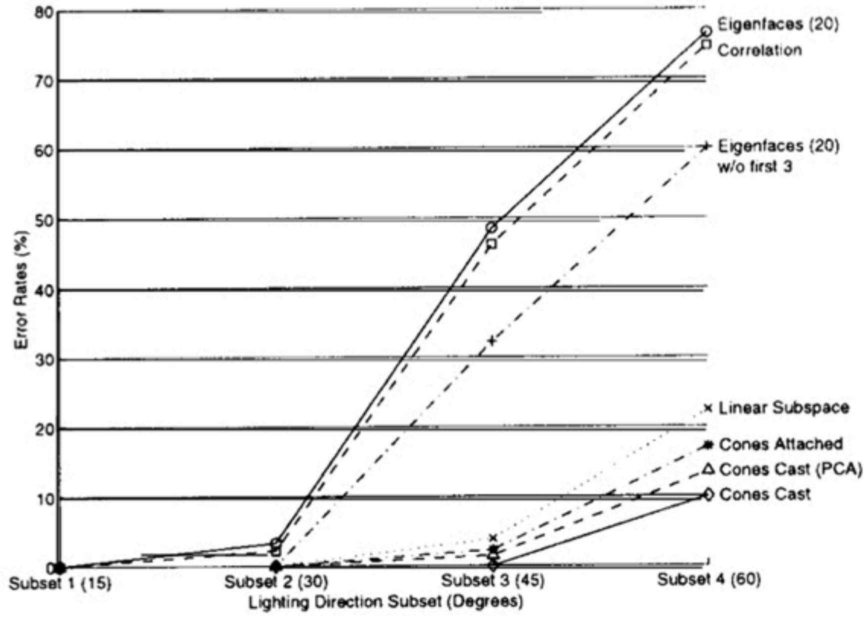

Figure 3 **(Georghiades et al., 58)**

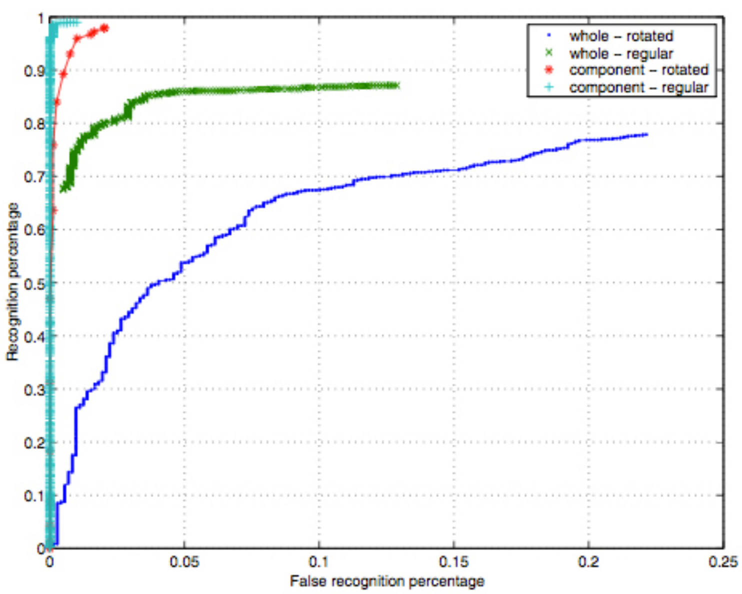

Figure 4
ROC curve of component-based recognition system verses whole-face recognition system. Component-based system clearly outperforms whole face with surprising accuracy. (Huang et al., 340)

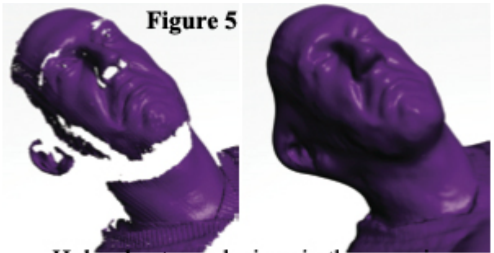

Figure 5

Holes due to occlusions in the scanning process (left) are filled by the implicit surface (right).
(Steinke et al., 290)

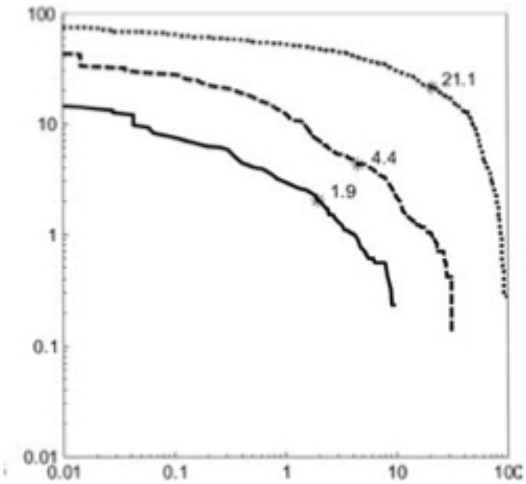

Figure 6

ROC curve. (Near) Isometric facial recognition outperforms 3D rigid facial surface matching and 2D eigenface algorithms.
(Bronstein et al., 24)

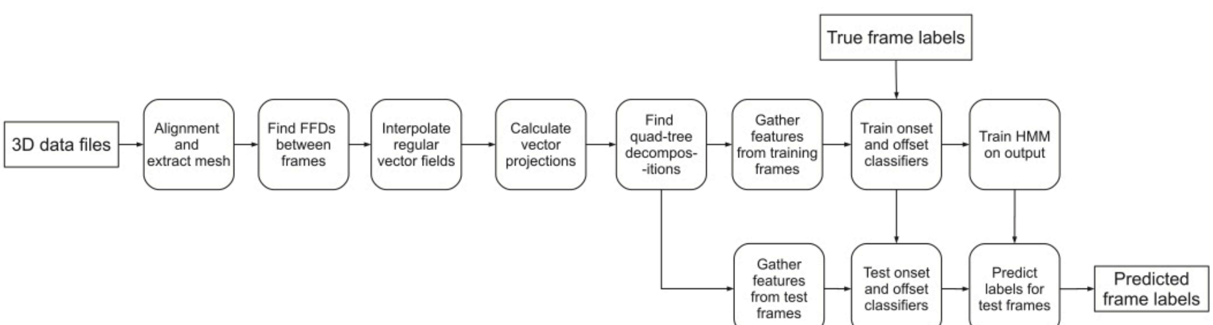

Figure 7
An overview of the fully automated system including motion caption, feature extraction, classification, and training and testing. (Sandbach et al., 4)

5 - SOURCES CITED

Belhumeur, P. N., Hespanha, J. P., and Kriegman, D. J. *Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997. 19, 711-720.
(Belhumeur et al.)

Belhumeur, P. N. and Kriegman, D. J. *What is the Set of Images of an Object Under All Possible Illumination Conditions?* IEEE Proc. Conf. Computer Vision and Pattern Recognition, 1996. Appearing in: International Journal of Computer Vision, 1998. 28, 245-260.
(Belhumeur and Kriegman)

Bowyer, K. W., Chang, K. I., Flynn, P. J., and Chen, X. *Face Recognition Using 2-D, 3-D, and Infrared: Is Multimodal Better Than Multisample?* Proceedings of the IEEE, 2006. 94, Pages 2000-2012.
(Bowyer et al.)

Bronstein, A. M., Bronstein, M. M., Kimmel, R. *Three-Dimensional Face Recognition*. International Journal of Computer Vision, 2005. 64, 5-30.
(Bronstein et al.)

G. Sandbach, et al., A dynamic approach to the recognition of 3D facial expressions and their temporal models, Image Vis. Comput. (2012), doi:10.1016/j.imavis.2012.01.006
(Sandbach et al.)

Galton, Sir Francis. *Personal Identification and Description*[2]. Nature, 1888. 38, 173-177.
(Galton, Sir F.)

Georghiades, A. S., Kriegman, D. J., and Belhumeur, P. N. *Illumination Cones for Recognition Under Variable Lighting: Faces*. IEEE Proc. Conf. Computer Vision and Pattern Recognition, 1998. 52-58.
(Georghiades et al.)

Gökberk, B., Salah, A. A., Akarun, L., Ethever, R., Riccio, D., and Dugelay, J-L. "3D Face Recognition." *Guide to Biometric Reference Systems and Performance Evaluation*. Ed. Petrovska-Delacrétaz, D., Chollet, G., and Dorizzi, B. London: Springer, 2009. 263-295.
(Gökberk et al.)

Heeger, D. *Signal Detection Theory*. Department of Psychology, New York University. 2003-2007, 1-10. URL: http://www.cns.nyu.edu/~david/handouts/sdt/sdt.html
(Heeger 2003-2007)

Huang, J., Blanz, V., and Heisele, B. “Face Recognition Using Component-Based SVM Classification and Morphable Models.” *Lecture Notes in Computer Science*. Berlin: Springer, 2002. 334-341.
(Huang et al.)

Kender, J. R. Class Notes. Visual Databases COMS E6735. Columbia University. Seeley W. Mudd, 535. February 16, 2012.
(Kender 2-16-12)

Lu, X., and Jain, A. K., *Deformation Modeling for Robust 3D Face Matching*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008. 30, 1346-1356.
(Lu and Jain)

Murase, H. and Nayar, S. K. *Visual Learning and Recognition of 3-D Objects from Appearance*. International Journal of Computer Vision, 1995. 14, 5-24.
(Murase and Nayar)

Pentland, A., Moghaddam, B., and Starner, T. *View-based and modular eigenspaces for face recognition*. Computer Vision and Pattern Recognition, 1994. 1994 IEEE Computer Society Conference. 1-7.
(Pentland et al.)

Poggio, T. and Edelman, S. “A network that learns to recognize three-dimensional objects.” Nature, 1990. 343, 263-266.
(Poggio and Edelman)

Quiroga, Q. R., Reddy, L., Kreiman, G., Koch, C., and Fried, I. *Invariant visual representation by single neurons in the human brain*. Nature, 2005. 435, 1102-1107.
(Quiroga et al.)

Reddy, L. and Kanwisher, N. *Coding of visual objects in the ventral stream*. Current Opinion in Neurobiology, 2006. 16, 408-414.
(Reddy and Kanwisher)

Steinke, F., Schölkopf, B., and Blanz, V. *Support Vector Machines for 3D Shape Processing*. Eurographics, 2005. 24, 285-294.
(Steinke et al.)

Tsao, D. Y. and Freiwald, W. A. *What’s so special about the average face?* Trends in Cognitive Sciences, 2006. 10, 391-393.
(Tsao and Freiwald)

Turk, M. and Pentland, A. *Eigenfaces for Recognition*. Journal of Cognitive Neuroscience, 1991. 3, 71-86.
(Turk and Pentland)

Zhao, W., Chellappa, R., Phillips, P. J., and Rosenfeld, A. *Face Recognition: A Literature Survey*. ACM Computing Surveys, 2003. 35, 399-458.
(Zhao et al.)