

An Introduction to Biometric-completeness: The Equivalence of Matching and Quality

P. Jonathon Phillips

J. Ross Beveridge

Abstract—This paper introduces the concept of **biometric-completeness**. A problem is **biometric-complete** if solving the problem is “equivalent” to solving a biometric recognition problem. The concept of biometric-completeness is modeled on the informal concept of artificial intelligence (AI) completeness. The concept of biometric-completeness is illustrated by showing a formal equivalence between biometric recognition and quality assessment of biometric samples. The model allows for the inclusion of quality of biometric samples in verification decisions. The model includes most methods for incorporating quality into biometric systems. The key result in this paper shows that finding the perfect quality measure for any algorithm is equivalent to finding the perfect verification algorithm. Two results that follow from the main result are: finding the perfect quality measure is equivalent to solving the open-set and closed-set identification problems; and that a universal perfect quality measure cannot exist.

I. INTRODUCTION

When attempting to solve computer science or pattern recognition problems, one should keep mind the Rolling Stones’ lyric “You can’t always get what you want. But if you try sometime ... You just might find you get what you need!”¹ Many times optimal solutions do not exist or proposed short cuts to solving the problem are as difficult to solve as the original problem. There is a rich history of characterizing the fundamental difficulty of problems in mathematics, computer science, artificial

intelligence, and machine learning. The continuum hypothesis, Godel’s incompleteness theorem, and the halting problem are examples where theory formally bounds an hypothesis that can be proved and a problem that can be solved. In algorithm theory, establishing an equivalence between one class of problem and another with known difficulty plays a particularly key role in characterizing problem difficulty, with NP-completeness being a prime example. Taking inspiration from the concept of NP-completeness, which is a formal mathematical concept, the artificial intelligence (AI) community has introduced the idea AI-completeness [9]. A problem is AI-complete if solving this problem is equivalent to solving the general artificial intelligence problem.

Characterizing the difficulty of a problem, at a theoretical level, as being very hard does not imply that research should cease. Rather it places the difficulty of a problem in perspective. Knowing a problem is NP-complete can redirect research in more fruitful directions. One example is the traveling salesperson problem, an NP-complete problem. A rich literature has emerged focused on approximation algorithms [1]. Another NP-complete example is the 3-satisfiability problem. In this problem there has been substantial work in characterizing tractable instances of the problem. In artificial intelligence, having a feel for the landscape of problem difficulty helps to set realistic goals for potential solutions.

Inspired by the notion of AI-completeness, we introduce the concept of *biometric-completeness*. A problem is biometric-complete if solving the problem is “equivalent” to solving the general biometric recognition problem. We illustrate the concept of biometric-completeness by presenting a

P. Jonathon Phillips is with the National Institute of Standards and Technology, Gaithersburg, MD 20899 jonathon@nist.gov

J. Ross Beveridge is with the Department of Computer Science, ColoradoStateUniversity, FortCollins, CO 80523 ross@cs.colostate.edu

¹From the song “You Can’t Always Get What You Want” by M. Jagger and K. Richards on the Rolling Stones’ album “Let It Bleed,” 1969.

formal equivalence of matching biometric samples and characterizing the quality of biometric samples.

In biometrics, one current avenue of research is to develop quality measures for biometric samples. The goal of a quality measures is to be able to identify biometric samples that increase the accuracy of a biometric recognition algorithm. In this paper, we present a model for incorporating quality scores into the biometric matching problem; from this model we show that finding the perfect quality measure is equivalent to finding a perfect biometric recognition algorithm; and that there does not exist a perfect universal quality measure.

This equivalence result should no more discourage the study of alternative quality measures than NP-completeness should discourage work on the traveling salesperson problem. What the result does accomplish is to allow researchers to focus on what is of practical value and avoid becoming entangled in arguments about universal and complete quality measures.

II. OVERVIEW OF MAIN RESULTS

A general model is outlined for the inclusion of quality of biometric samples in verification decisions, Section IV formally defines this model. The model includes most methods for incorporating quality into biometric systems. The key result in this paper shows that finding the perfect quality measure for any algorithm is equivalent to finding the perfect verification algorithm. In Section VI the proofs of the equivalence results are given. Two results that follow from the main result are: finding the perfect quality measure is equivalent to solving the open-set and closed-set identification problems; and there cannot exist a universal perfect quality measure.

The basic model for recognition is one-to-one verification. In one-to-one verification, the input to an algorithm are two biometric samples and the algorithm decides if the biometric samples come from the same individual. The results generalize to one-to-many verification, and open and closed identification (open-set identification is also known as the watch-list problem).

In Grother and Tabassi [6], a quality metric is

defined as being able to predict performance of a biometric sample. Their work introduced the NIST fingerprint image quality metric (NFIQ). In NFIQ, the quality for fingerprint is measured by a single value and a higher quality value corresponds to greater recognition rates.

Concurrent to the biometric quality work, was research in characterizing the performance of face recognition algorithms in terms of covariates [2], [3], [5]. Covariates could be subject, image, or quality. Subject covariates are attributes of the person being recognized, such as age, gender or race. Subject covariates can be transitive properties of subjects, such as smiling or wearing glasses. Image covariates are attributes of the image or sensor, such as size of the face or focus of the camera. A quality measure is a covariate that is measurable, predictive of performance and actionable. A quality measure can be a subject or images covariate such as focus, size of the face, or expression.

There are two key findings from covariate analysis of face recognition algorithms that are addressed in our model. The first is that multiple covariates can characterize the quality of an image. Therefore, in the model in this paper, quality can be characterized by multiple values or quality measures. The second is that performance depends on the quality of both the biometric samples being matched. An example is expression in face recognition. It is significantly better to match either a smiling face to a smiling face or a neutral to a neutral, than to match across a facial expression [2]. Our model includes the ability to characterized quality as an interaction between the pairs of biometric samples being compared.

One common method for reporting performance is to assess the quality metrics ability to predict performance on a set of images [3], [4], [6]. Performance on higher quality biometric samples should be better than performance on lower quality samples. These methods characterize performance over a set of biometric samples. However, in our model, we adopt the approach of characterizing quality performance at the level of individual similarity scores (a comparison between two biometric samples). A quality measure's performance is characterized by

its ability to predict when an algorithm will fail or succeed in verifying two biometric samples.

A perfect biometric algorithm does not make any errors. A perfect quality measure always accurately reports when an algorithm will make a mistake. Given any algorithm and a perfect quality measure, a perfect algorithm can be constructed. If the perfect quality measure estimates that the algorithm will make the wrong decision, then reverse the original decision, otherwise, don't change the original decision. Thus, solving the quality problem solves the biometric recognition problem. This means the quality and recognition problems are equivalent and the quality problem is biometric-complete.

A perfect quality measure is tuned to the errors of an individual algorithm. Two algorithms will have two different sets of errors and therefore separate perfect quality measures. Thus, it is not possible to have a universal perfect quality measure.

III. BIOMETRIC-COMPLETENESS

Inspired by the definition of AI-complete, a definition for biometric-complete will be introduced. AI-complete is not a mathematically formal definition. Rather it summarizes the belief of the AI community that there is a class of hard AI problems—solving one of these problems, solves all the problems. Shapiro [9] defines AI-complete as

Several of these broad areas can be considered *AI-complete*, in the sense that solving the problem of the area is equivalent to solving the entire AI problem—producing a generally intelligent computer program.

The first step is to define what is meant as *the biometric problem*—there could be different definitions depending on the context. “The” biometric problem is to develop a perfect or highly accurate biometric recognition system. “The” is quoted to emphasize that there could be varying definitions of “the” biometric problem depending on context. For example, the biometric problem associated with iris recognition is significantly different than the biometric problem associated with unconstrained face recognition. The concept of biometric-complete is defined to be:

A problem area is considered to be *biometric-complete* if a solution in this area is equivalent to solving the core biometric problem of uniquely identifying people from biometric samples, i.e. the biometric recognition problem.

Problem areas in general represent any task defined relative to biometrics. Active research areas include efforts to quantify biometric quality, usability of biometrics, and sensor design. This paper will consider specifically efforts to define biometric quality scores. Sections IV, V and VI formally shows that solving the quality problem is equivalent to finding a perfect recognition algorithm. Thus, the quality problem is also biometric-complete.

A related problem is biometric usability, where biometric usability is defined as the task of developing user interfaces that allow for the consistent collection of high quality samples from people. Thus, the biometric usability problem is similar to the biometric quality problem and we argue that solving the usability problem is biometric-complete.

Quality and usability are examples of biometric-complete problems that need to be addressed by the majority of biometric systems. Because solving either the quality or usability problems is biometric-complete, finding approximate solutions to both has the potential to substantially increase system performance. The characterization of a set of problems as biometric-complete, suggests that finding a series of approximate solutions is preferred over searching for an optimal solution.

IV. MODEL

The model in this paper for biometric recognition consists of two parts: algorithms and quality functions. The definition of the model proceeds by first defining the properties of algorithms. This is followed by defining quality functions.

The space of target samples will be denoted by $\mathcal{I}_T$ and the space of query samples will be denoted by $\mathcal{I}_Q$. An algorithm A is a function $A : \mathcal{I}_T \times \mathcal{I}_Q \rightarrow \{0, 1\}$, where 1 means that the target and query samples are from the same person and 0 means the samples are from different people.

In the model in this paper, the definition of an algorithm is different than in the biometrics evaluation literature. First, algorithms in the model are restricted to one-to-one verification, and techniques such as gallery normalization for cohort normalization are not included in the model. Second, algorithms are generally modeled as $A_S : \mathcal{I}_T \times \mathcal{I}_Q \rightarrow \mathbb{R}$, where $A_S \equiv s(t, q)$ and $s(t, q)$ is a similarity score. Without loss of generality, a large similarity score implies that two biometric samples are more likely to be from the same person. An algorithm that produces similarity score can be converted to a family of algorithms A_τ , for a threshold τ . $A_\tau = 0$ if $s(t, q) < \tau$ and $A_\tau = 1$ if $s(t, q) \geq \tau$. The results of this paper apply to both gallery normalization verification algorithms and algorithms that produce a similarity score.

In a typical scenario for quality metrics, first, a biometric is acquired. Second, the quality score is computed. Third, if the quality score meets a criteria, the biometric sample is kept. In this scenario, there are quality functions for target and query samples denoted by r_T and r_Q . The quality function for target is $r_T : \mathcal{I}_T \rightarrow \mathcal{X}_T$. The space $\mathcal{X}_T$ is a general space. It can be an ordered or unordered set or a vector space such as $\mathbb{R}^n$. More generally, $\mathcal{X}_T$ can be an n-tuple of components that are a combination of ordered and unordered sets, and $\mathbb{R}$. The origin of a n-tuple could be a covariate analysis [2], [3], [5]. A biometric sample t is of sufficient quality if $r_T(t) \in \mathcal{X}_T^*$, where $\mathcal{X}_T^* \subset \mathcal{X}_T$ is the region of high quality samples. Quality functions r_Q for query samples are similarly defined along with $\mathcal{X}_Q^*$ and $\mathcal{X}_Q$. It is possible that $r_T(t)$ (resp. $r_Q(t)$) is the identity function, and $\mathcal{X}_T = \mathcal{I}_T$ and $\mathcal{X}_T^* \subset \mathcal{I}_T$ (resp. $\mathcal{X}_Q = \mathcal{I}_Q$ and $\mathcal{X}_Q^* \subset \mathcal{I}_Q$). Quality functions are defined separately for target and query samples because there might be different quality standards for enrollment and acquisition biometric samples.

The above quality regime only allows an algorithm A to compare a target sample t and a query sample q if $r_T(t) \in \mathcal{X}_T^*$ and $r_Q(q) \in \mathcal{X}_Q^*$. Here the quality scores are computed independently for the target and query samples. This regime does not allow for explicit interaction between target and query samples. However, it has been shown that

there are interactions between quality covariates for target and query samples that effect performance [2], [3]. Returning to the example in our introduction, for faces it is better to match smiling face images or neutral face images than match across face expression [2].

To allow for interactions between target and query samples that effect performance, a quality function r is introduced. The quality function is $r(r_T(t), r_Q(t)) : \mathcal{X}_T \times \mathcal{X}_Q \rightarrow \{0, 1\}$, where 1 means that the sample pair (t, q) are of sufficient quality, and 0 means the pair are not of sufficient quality. For a concise notation, $r(t, q)$ will refer to $r(r_T(t), r_Q(t))$.

V. PERFORMANCE OF A QUALITY METRIC

The traditional method for measuring the effectiveness of a quality function is its ability to predict performance [2], [3], [4], [6]. The strength of this method is that the effectiveness of a quality function is measured with respect to system performance. In the model in this paper, it is necessary to look directly at the relationship between the quality function and algorithm performance. Performance of a quality function is directly measured by its effectiveness in predicting when an algorithm fails, this method is also used in Li et al [7] and Scheirer and Boulton [8].

We will now proceed with a formal definition. Let $g : \mathcal{I}_T \times \mathcal{I}_Q \rightarrow \{0, 1\}$ be the ground truth functions, with $g(t, q) = 0$ meaning that t and q are samples of different people, i.e. t and q are a non-match pair. When $g(t, q) = 1$, the samples t and q are from the same person, i.e. t and q are a match pair. The accuracy function $e_A : \mathcal{I}_T \times \mathcal{I}_Q \rightarrow \{0, 1\}$ for algorithm A tells whether algorithm A correctly classified the pair of samples (t, q) . When the accuracy function $e_A(t, q) = 0$ it means the algorithm was incorrect and when $e_A(t, q) = 1$ it means the algorithm was correct. The region of $\mathcal{I}_T \times \mathcal{I}_Q$ where $e_A(t, q) = 0$ will be denoted by E_0 . This is the region of $\mathcal{I}_T \times \mathcal{I}_Q$ that an algorithm A gives the wrong answer. The region of $\mathcal{I}_T \times \mathcal{I}_Q$ where A gives the correct answer is similarly defined and is denoted by E_1 . Traditional measures of verification performance can be computed from the algorithm

accuracy function. The false reject (FRR) and false accept (FAR) rates are

$$\text{FRR} = \frac{\int_{\Omega_1} 1 - e_A(t, q) \, d\mathcal{I}_T \, d\mathcal{I}_Q}{\int_{\Omega_1} d\mathcal{I}_T \, d\mathcal{I}_Q} \quad (1)$$

and

$$\text{FAR} = \frac{\int_{\Omega_0} 1 - e_A(t, q) \, d\mathcal{I}_T \, d\mathcal{I}_Q}{\int_{\Omega_0} d\mathcal{I}_T \, d\mathcal{I}_Q}. \quad (2)$$

The accuracy of a quality function $r(t, q)$ is related to the accuracy of an algorithm A . A quality function $r(t, q)$ should predict when an algorithm A will make a classification error, which occurs when $e_A(t, q) = 0$. Thus, for algorithm A , the perfect quality function $r^*(t, q) \equiv e_A(t, q)$.

The performance of a quality function can be characterized as a signal detection problem with detection rate, P_d , and false alarm rate, F_a . A correct detection occurs when $e_A(t, q) = 1$ and $r(t, q) = 1$, and a false alarm occurs when $e_A(t, q) = 0$ and $r(t, q) = 1$. These rates, or probabilities are

$$P_d = 1 - \frac{\int_{E_1} e_A(t, q) \text{ XOR } r(t, q) \, d\mathcal{I}_T \, d\mathcal{I}_Q}{\int_{E_1} d\mathcal{I}_T \, d\mathcal{I}_Q} \quad (3)$$

and

$$F_a = \frac{\int_{E_0} e_A(t, q) \text{ XOR } r(t, q) \, d\mathcal{I}_T \, d\mathcal{I}_Q}{\int_{E_0} d\mathcal{I}_T \, d\mathcal{I}_Q}. \quad (4)$$

VI. EQUIVALENCE RESULTS

An algorithm A is perfect if it does not make any classification errors; e.g, its FAR = 0, and its FRR = 0. Similarly, a quality function is perfect if it does not make any classification errors, which implies $r(t, q) = e_A(t, q)$. The following theorem states the basic equivalence theorem between perfect one-to-one verification algorithms and perfect quality functions.

Theorem 1. *For a one-to-one verification algorithm $A(t, q)$, finding the a perfect quality function $r^*(t, q)$ is equivalent to finding a perfect one-to-one verification algorithm.*

Proof: For an algorithm A and its perfect quality function $r^*(t, q)$, a new algorithm A^* is defined as

$$A^*(t, q) = \begin{cases} A(t, q) & \text{if } r^*(t, q) = 1 \\ A(t, q)^c & \text{if } r^*(t, q) = 0 \end{cases}, \quad (5)$$

where c is the complementary function. The algorithm A^* is perfect.

Corollary 2. *Theorem 1 holds for an algorithm A_S that produces a similarity score $s(t, q)$.*

Proof: We define a new verification algorithm $A_\tau(t, q)$ which is equal to 0 if $s(t, q) < \tau$ and equal to 1 if $s(t, q) \geq \tau$, where τ can arbitrarily chosen in the range of A_S ; i.e., $\inf s(t, q) < \tau < \sup s(t, q)$.

Theorem 3. *Finding a perfect quality function $r^*(t, q)$ for a one-to-one verification algorithm $A(t, q)$, is equivalent to finding a perfect open-set identification algorithm.*

Proof: Construct the perfect algorithm A^* according to Eq. (5). The gallery in an open-set identification problem consists of n samples $\{t_1, \dots, t_n\}$ of n different individuals. Next reduce the open-set query to n one-to-one verification problems $\{A^*(t_1, q), \dots, A^*(t_n, q)\}$. Since the results of each of the one-to-one verification queries is correct, the answer to the open-set query is correct.

VII. UNIVERSALITY RESULTS

One of the goals of developing quality functions is for them to be “universal.” In practical terms, one is only interesting in determining if a set of quality functions is effective for the best performing algorithms that are currently available. Universality of quality functions will be addressed at a slightly more abstract level. Universality will be restricted to a set of algorithms $A_1, \dots, A_n$. This set can be a set of better performing algorithms, a set of related algorithms, or an arbitrary set of algorithms. In this paper we will restrict our attention to the existence of a perfect quality function for all the algorithms in the set $A_1, \dots, A_n$. We will first look at the case for only two algorithms. The key results looks at the case when there are only two algorithms.

Theorem 4. *Given two algorithms A_1 and A_2 , there does not exist a quality function r^{**} that is perfect for both algorithms A_1 and A_2 .*

Proof: Let $I_{A_i} \equiv I_{[A_i(t,q)=1]}(t, q)$, which is the region of $\mathcal{I}_T \times \mathcal{I}_Q$ where the algorithm A_i is correct. Without loss of generality, $D = I_{A_1} \cap I_{A_2}^c \neq \emptyset$, otherwise chose $D = I_{A_1}^c \cap I_{A_2}$. Let r^{**} be a perfect quality functions for A_1 and A_2 . Now, for r^{**} to be a perfect quality function for algorithm A_1 , $r^{**}|_D = 1$; and for r^{**} to be a perfect quality function for algorithm A_2 , $r^{**}|_D = 0$, which is a contraction.

The obvious corollary to Theorem 4 is

Corollary 5. *Given a set of algorithms $A_1, \dots, A_n$, there does not exist a quality function r^{**} that is perfect for the algorithms $A_1, \dots, A_n$.*

VIII. CONCLUSION

The concept of biometric-completeness was introduced as a method for describing which biometric problems are equivalent to solving the biometric recognition problem. The idea Biometric-completeness was formally defined for the quality problem. The biometric quality problem was shown to be equivalent to solving the biometric recognition problem. Thus, the quality problem is biometric-complete. Biometric-completeness provides a framework for understanding the fundamental difficulty of biometric problems. This understanding provides guidance for allocating resources for solving biometric problems.

IX. ACKNOWLEDGMENTS

PJP acknowledges the support of the the Biometric Task Force, the Department of Homeland Security's Directorate for Science and Technology, the Intelligence Advanced Research Projects Activity (IARPA), the Federal Bureau of Investigation (FBI), and the Technical Support Working Group (TSWG). JRB was funded in part by the Technical Support Working Group (TSWG) under Task T-1840C. The identification of any commercial product or trade name does not imply endorsement or recommendation by the Colorado State University or the National Institute of Standards and Technology.

REFERENCES

- [1] D. Applegate, R. Bixby, V. Chvátal, and W. Cook. *The Traveling Salesman Problem: A Computational Study*. Princeton University Press, 2006.
- [2] J. R. Beveridge, G. H. Givens, P. J. Phillips, and B. A. Draper. Factors that influence algorithm performance in the Face Recognition Grand Challenge. *Computer Vision and Image Understanding*, (in press).
- [3] J. R. Beveridge, G. H. Givens, P. J. Phillips, B. A. Draper, and Y. M. Lui. Focus on quality, predicting FRVT 2006 performance. In *Eighth International Conference on Automatic Face and Gesture Recognition*, 2008.
- [4] P. J. Flynn and P. J. Phillips. ICE mining: Quality and demographic investigation of ICE 2006 performance results. Technical report, National Institute of Standards and Technology, 2008.
- [5] G. H. Givens, J. R. Beveridge, B. A. Draper, P. J. Grother, and P. J. Phillips. How features of the human face affect recognition: a statistical comparison of three face recognition algorithms. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 381–388, 2004.
- [6] P. Grother and E. Tabassi. Performance of biometric quality measures. *IEEE Trans. PAMI*, 29:531–543, 2007.
- [7] W. Li, X. Gao, and T. Boulton. Predicting biometric system failure. In *Proceedings of the 2005 IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety*, pages 57–64, 2005.
- [8] W. Scheirer and T. E. Boulton. A fusion-based approach to enhancing multi-modal biometric recognition system failure prediction and overall performance. In *2nd IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 1–7, 2008.
- [9] S. C. Shapiro. Artificial intelligence. In S. C. Shapiro, editor, *Encyclopedia of Artificial Intelligence, Second Edition*, pages 54–57. John Wiley & Sons, Inc., New York, 1992.