Search Pruning in Video Surveillance Systems: Efficiency-Reliability Tradeoff

Antitza Dantcheva, Arun Singh, Petros Elia, Jean-Luc Dugelay
EURECOM
Sophia Antipolis, France
{Antitza.Dantcheva, Arun.Singh, Petros.Elia, Jean-Luc.Dugelay}@eurecom.fr

Abstract

In the setting of computer vision, algorithmic searches often aim to identify an object of interest inside large sets of images or videos. Towards reducing the often astronomical complexity of this search, one can use pruning to filter out objects that are sufficiently distinct from the object of interest, thus resulting in a pruning gain *of an overall reduced search space.*

Motivated by practical computer vision based scenarios such as time-constrained human identification in biometric-based video surveillance systems, we analyze the stochastic behavior of time-restricted search pruning, over large and unstructured data sets which are furthermore random and varying, and where in addition, pruning itself is not fully reliable but is instead prone to errors. In this stochastic setting we apply the information theoretic method of types as well as information divergence techniques to explore the natural tradeoff that appears between pruning gain and reliability, and proceed to study the typical and atypical gain-reliability behavior, giving insight on how often pruning might fail to substantially reduce the search space. The result, as is, applies to a plethora of computer vision based applications where efficiency and reliability are intertwined bottlenecks in the overall system performance, and the simplicity of the obtained expressions allows for rigorous and insightful assessment of the pruning gain-reliability behavior in such applications, as well as for intuition into designing general object recognition systems.

1. Introduction

In recent years we have experienced an increasing need to structure and organize an exponentially expanding volume of data that may take the form of, among other things, images and videos. Crucial to this effort is the often computationally expensive task of algorithmic search for specific elements placed at unknown locations inside large data sets. To limit computational cost, pre-filtering such as pruning can be used, to quickly eliminate a portion of the initial data, an action which is then followed by a more precise and complex search within the smaller subset of the remaining data. Such pruning methods can substantially speed up the search, at the risk though of missing the target, thus reducing the overall reliability. Common pre-filtering methods include video indexing and image classification with respect to color [1], patterns, objects [2], or feature vectors [3].

1.1. Categorization-based pruning of time-constrained searches over error-inducing stochastic environments

Our interest in analyzing the efficiency vs. reliability tradeoff, focuses on the realistic setting where the search is time-constrained and where, as we will see later on, the environment in which the search takes place is stochastic, dynamically changing, and can cause search errors. We note here that there is a fundamental difference between search in unstructured versus structured data, where the latter can be handled with very efficient algorithms, such as different search and bound algorithms (cf. [12]). One widely known practical scenario that adheres to the above stochastic setting, is the scenario of biometric-based video surveillance. In this setting it is of interest to locate/retrieve a target subject at a specific time and location out of all subjects present in the specific setting. In this scenario, a set of subjects can be pruned by means of categorization that is based on different combinations of soft biometric traits such as gender, height or hair color. The need for such biometrically-based search pruning is often brought to the fore in cases such as in the 2005 London bombing where a sizeable fraction of the police force worked for days to screen a subset of the available surveillance videos relating to the event.

1.2. On the adopted approach

The analysis here, despite being often presented in the language of video surveillance, can be generally applied to all domains of computer vision that naturally adhere to the setting of categorization-based pruning in time-constrained searches over error-prone stochastic environments. In such general settings, pruning of large datasets, be they images or video, involves a corresponding natural tradeoff between pruning efficiency[1] and reliability. We here concisely describe this tradeoff, for a very general setting of statistics and algorithmic capabilities.

We clearly note that the work here *is not* meant to model

[1] We note that the terms efficiency and gain are used interchangeably.

the population statistics or the algorithmic properties. Instead the presented analysis holds for a general class of statistics and classification algorithms. This analysis takes as input the aforementioned models, and provides as output the (average and tail behavior of the) different efficiency and reliability capabilities.

Towards concisely and meaningfully handling the often intractable complexity of the problem, we employ the powerful mathematical machinery of the *method of types* (cf. [10]) underlying information theory, which has played a pivotal role in resolving similar efficiency vs. reliability problems in several other domains such as in network communications [11]. In conjunction with information divergence techniques, the above machinery is applied here to also provide a concise representation of the efficiency-vs-reliability problem in search pruning of random and large databases, which is one of the main problems encountered in several computer vision applications.

We note that the application of this machinery assumes the presence of large databases. We argue that this assumption is not only acceptable but is in fact necessary, given the massive data volumes (resp. the massive populations) that pruning applications (resp. surveillance systems) must efficiently handle. As the size increases, the presented asymptotic results not only become more precise but they also become more useful, mainly because simulating the behavior of massive systems can often be computationally prohibitive.

Finally this same approach allows us to cover a broad spectrum of system behavior, spanning the entire range between average-case and worst-case behaviors. This approach is particularly suited for scenarios where system failure, even when rare, carries a non-negligible cost. This same approach is necessary also because it is often the case that, in the presence of sizeable databases, average-case analysis may be too optimistic and may fail to capture the core of the problem, and on the other hand worst-case analysis is often overly pessimistic and thus may again fail to give proper insight.

1.3. Categorization based pruning

We consider the scenario where we search for a specific *object*[2] *of interest,* denoted as v', belonging to a large and randomly drawn *authentication group* $\boldsymbol{v}$ of n objects, where each object belongs to one of ρ *categories*. The elements of the set (authentication group) $\boldsymbol{v}$ are derived randomly from a larger population, which adheres to a set of population statistics. A category corresponds to objects that adhere to a specific combination of characteristics, so for example in the setting of surveillance systems, one may consider a category consisting of blond, tall, females.

With n being potentially large, we seek to simplify the search for object v' within $\boldsymbol{v}$ by *algorithmic pruning* based on categorization, i.e., by first identifying the objects that potentially belong to the same category as v', and by then pruning out all other objects that have not been estimated to share the same traits as v'. Pruning is then expected to be followed by careful search of the remaining unpruned set. Such categorization-based pruning allows for a search speedup through a reduction in the search space, from $\boldsymbol{v}$ to some smaller and easier to handle set $\mathcal{S}$ which is the subset of $\boldsymbol{v}$ that remains after pruning, cf. Fig. 1. This reduction though happens in the presence of a set of *categorization error probabilities* $\{\epsilon_f\}$, also referred to as *confusion probabilities*[3] (cf. [16, 15]), that essentially describe how easy it is for categories to be confused, hence also describing the probability that the estimation algorithm erroneously prunes out the object of interest, by falsely categorizing it. This confusion set, together with the set of population statistics $\{p_f\}_{f=1}^{\rho}$ which describes how common a certain category is inside the large population, jointly define the statistical behavior of search pruning, which we will explore. The above aspects will be precisely described later on.

Example 1 Pruning for video surveillance

a) Categorization pruning with soft biometrics *In the setting of surveillance systems, categorization based pruning is achieved using soft biometric identifiers which are human physical, behavioral or adhered characteristics, which carry information about the individual, are computationally efficient, easy to acquire, but which are generally not sufficient to fully authenticate an individual (cf. [4, 5, 14, 13]). Scientific work on using soft biometrics for pruning the search can be found in [6, 7], where a multitude of attributes, like age, gender, hair and skin color were used for classification of a face database, as well as in [8, 9] where the impact of pruning traits like age, gender and race was identified in enhancing the performance of regular biometric systems. We henceforth refer to a system which extracts features and classifies them in pre-defined categories, as a soft biometric system (SBS).*

b) Pruning the search in surveillance systems *An example of a sufficiently large population includes the inhabitants of a certain city, and an example of a randomly chosen authentication group (n-tuple) $\boldsymbol{v}$ includes the set of people captured by a video surveillance system in the aforementioned city between 11:00 and 11:05 yesterday. An example SBS could be able to classify 5 instances of hair color, 6 instances of height and 2 of gender, thus being able to differentiate between $\rho = 5 \cdot 6 \cdot 2 = 60$ distinct categories. An example search could seek for a subject that was described to belong to the first category of say, blond and tall females. The subject and the rest of the authentication group of $n = 10000$ people, were captured by a video-surveillance system at approximately the same time at a specific part of the city. In this city, each SBS-based category appears with probability $p_1, \cdots, p_{60}$, and each such category can be confused for the first category with probability $\epsilon_2, \cdots, \epsilon_{60}$. The SBS makes an error whenever v' is pruned out, thus it allows for reliability of ϵ_1. To clarify, having $p_1 = 0.1$ implies that approximately one in ten city*

[2] The terms object and subject are here used interchangeably.

[3] A formal definition of these probabilities will follow, where this definition will clarify the special role assigned to ϵ_1.

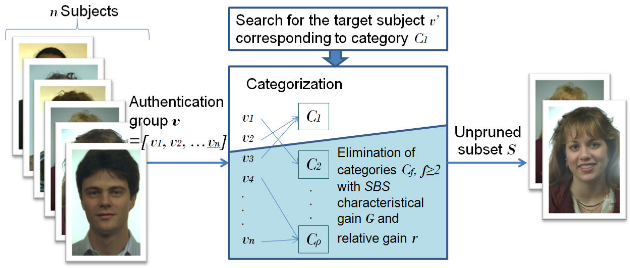

Figure 1. System overview. We clarify that this image is simply meant as a visual aid to the reader rather than a faithful representation of the image quality encountered in surveillance systems.

inhabitants are blond-tall-females, and having $\epsilon_2 = 0.05$ *means that the system (its feature estimation algorithms) tends to confuse the second category for the first category with probability equal to* 0.05 *(cf. Fig. 4).*

What becomes apparent is that a more aggressive pruning of subjects in $\boldsymbol{v}$ results in a smaller $\mathcal{S}$ and a higher pruning gain, but as categorization entails estimation errors, such a gain could come at the risk of erroneously pruning out the subject v' that we are searching for, thus reducing the system reliability.

Stochastic system behavior Reliability and pruning gain are naturally affected by, among other things, the distinctiveness and differentiability of the subject v' from the rest of the subjects in the specific authentication group $\boldsymbol{v}$ over which pruning will take place at a particular instance. In several scenarios though, this distinctiveness changes randomly because $\boldsymbol{v}$ itself changes randomly, as well as because the detection medium fluctuates. These introduce a stochastic environment. Consequently depending on the instance during which v' and its surroundings $\boldsymbol{v} - v'$ were captured by the system, $\boldsymbol{v}$ may consist of neighboring subjects that look similar to the subject of interest v', or $\boldsymbol{v}$ may consist of subjects that look sufficiently different from the subject of interest. Naturally the first case is generally expected to allow for a lower pruning gain than the second case.

The pruning gain and reliability behavior can also be affected by the system design. At one extreme we find a very conservative system that prunes out a member of $\boldsymbol{v}$ only if it is highly confident about its estimation and categorization, in which case the system yields maximal reliability (near-zero error probability) but with a much reduced pruning gain. At the other extreme, we find an effective but unreliable system which aggressively prunes out subjects in $\boldsymbol{v}$ without much regard for erroneous pruning, resulting in a potentially much reduced search space ($|\mathcal{S}| << n$), at a high risk though of an error. In the above, $|\mathcal{S}|$ denotes the cardinality of set $\mathcal{S}$, i.e., the number of subjects that were not pruned out.

1.4. Contributions

In the next section we elaborate on the concept of *pruning gain*, which describes, as a function of pruning reliability, the reduction of the set size after pruning: for example a pruning gain of 2 implies that pruning managed to halve the size of the original set. Section 3 applies the method of types and information divergence techniques to provide statistical analysis of the pruning gain as a function of reliability, offering insight on how often pruning fails to be sufficiently helpful, given a set of system resources. In the process we try to provide some intuition through examples on topics such as, how the system gain-reliability performance suffers with increasing confusability of categories, or on whether searching for a rare looking subject renders the search performance more sensitive to increases in confusability, than searching for common looking subjects. Subsection 3.1 draws from the asymptotic results and provides average case analysis of the gain, suggests the measure of average *goodput* as a measure that jointly maps the efficiency and reliability capabilities of a system, as well as provides several clarifying examples.

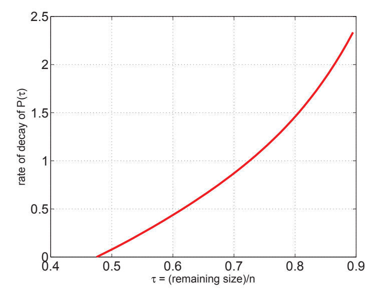

Figure 2. Asymptotic rate of decay of $P(|\mathcal{S}| > \tau n)$, for $\rho = 3$, reliability 0.8, population statistics $p_1 = 0.4, p_2 = 0.25, p_3 = 0.35$ and confusability parameters $\epsilon_2 = 0.2, \epsilon_3 = 0.3$.

Before proceeding with the analysis we hasten to give some insight, again in the language of video surveillance, as to what is to come. In the setting of large n, the results of Section 3 are better illustrated here with an example. In this we ask what the probability is that a CV-based pruning system that can identify $\rho = 3$ categories, that searches for a subject of the first category, that has 80 percent reliability ($\epsilon_1 = 0.8$), that introduces confusability parameters $\epsilon_2 = 0.2, \epsilon_3 = 0.3$ (cf. Fig 4) and that operates over a population with statistics $p_1 = 0.4, p_2 = 0.25, p_3 = 0.35$, will in fact prune the search to only a fraction of $\tau = |\mathcal{S}|/n$. We note that here τ is the inverse of the instantaneous pruning gain. We plot in Fig. 2 the asymptotic rate of decay for this probability,

$$J(\tau) := -\lim_{n\to\infty} \frac{\log}{n/\rho} P(|\mathcal{S}| > \tau n) \qquad (1)$$

for different values of τ. From the $J(\tau)$ in Fig. 2 we can draw different conclusions, such as:

- Focusing on $\tau = 0.475$ where $J(0.475) = 0$, we see that the size of the (after pruning) set $\mathcal{S}$ is typically

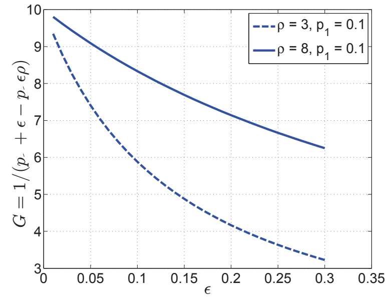

Figure 3. Average pruning gain $G = n/|\mathcal{S}|$, as a function of the confusability probability ϵ (for all categories), and for $p_1 = 0.1$. Plotted for $\rho = 3$ and $\rho = 8$.

(most commonly - with probability that does not vanish with n) 47.5% of the original size n. In the absence of errors, this would have been equal to $p_1 = 40\%$ (rather than 47.5%), but the errors cause a reduction of the average gain (from 47.5% to 40%).

- Focusing on $\tau = 0.72$, we note that the probability that pruning removes less than $1 - 0.72 = 28\%$ of the original set, is approximately given by $e^{-\rho n} = e^{-3n}$ ($J(0.72) \approx 1,\ \rho = 3$), whereas focusing on $\tau = 0.62$, we note that the probability that pruning removes less than $1 - 0.62 = 38\%$ of the original set, is approximately given by $e^{-\rho n/2} = e^{-3n/2}$ ($J(0.62) \approx 1/2$). The probability that pruning removes less than half the elements is approximately $P(\tau > 0.5) \approx e^{-3n/10}$ ($J(0.5) \approx 1/10$).

To clarify, the *rate function* $J(\tau)$ approximates $P(|\mathcal{S}| > \tau n)$ by describing its asymptotic rate of decay as n increases, in the sense that the plot of $-\log P(|\mathcal{S}| > \tau n)$ vs. n/ρ, after a sufficiently large n, would be a straight line with slope $J(\tau)$. Equivalently the plot of $\log P(|\mathcal{S}| > \tau n)$ vs. n would approximate a straight line with rate function $-\rho J(\tau)$, so in the above last example, the plot $\log P(\tau > 0.5)$ vs. n would, for sufficiently large n, be a straight line with slope $-3/10$. We finally note that as the population statistics change, so does J.

Moving to the special case of the average-case analysis, Subsection 3.1 directly tells us that the average pruning gain takes the form[4] of the inverse of $\sum_{f=1}^{\rho} p_f \epsilon_f$. This is illustrated in an example in Fig. 3 for varying confusability probability, for the case where the search is for an individual that belongs to the first category which occurs once every ten people ($p_1 = 0.1$), and for the case of two different systems that can respectively distinguish $\rho = 3$ or $\rho = 8$ categories. The expressions from the above graphs will be derived in detail later on.

[4]As a small sanity check we quickly note that applying the values of the p_i, ϵ_i from the example of Fig. 2 gives $\sum_{f=1}^{\rho} p_f \epsilon_f = 0.475$, which agrees with the above statement that the typical system behavior corresponds to $\tau = 0.475$ where $J(0.475) = 0$.

2. Gain v.s. reliability

As an intermediate measure of efficiency we consider the (instantaneous) *pruning gain*, defined here as

$$\mathcal{G}(\boldsymbol{v}) := \frac{n}{|\mathcal{S}|}, \tag{2}$$

which simply describes[5] the size reduction, namely from $\boldsymbol{v}$ to $\mathcal{S}$, and which can vary from 1 (no pruning gain) to n. In terms of system design, one could also consider the *relative gain*,

$$r(\boldsymbol{v}) := 1 - \frac{|\mathcal{S}|}{n} \in [0, 1], \tag{3}$$

describing the fraction of $\boldsymbol{v}$ that was pruned out.

It is noted here that $\mathcal{G}(\boldsymbol{v})$, and by extension $r(\boldsymbol{v})$, are random variables and randomly fluctuate as a function of, among other things, the relationship between $\boldsymbol{v}$ and v', the current estimation conditions as well as the error capabilities of the system. For example, as already implied before, we note that if $\boldsymbol{v}$ and v' are such that v' belongs in a category in which very few other members of $\boldsymbol{v}$ belong to, then pruning is expected to produce a very small $\mathcal{S}$ and a high gain. If though, at the same time, the estimation capabilities (algorithms and hardware) of the system result in the characteristics of v' being easily confusable with the characteristics of another populous category in $\boldsymbol{v}$, then $\mathcal{S}$ will be generally larger, and the gain smaller.

As a result, any reasonable analysis of the gain-reliability behavior must be of a statistical nature and must naturally reflect the categorization refinement, the corresponding estimation error capabilities of the system, as well as the statistics of the larger population. We proceed to clearly define these parameters.

2.1. Categorization, estimation capabilities and population statistics

2.1.1 Categorization and population statistics

In the setting of interest, we consider that $\boldsymbol{v}$ is chosen at random from a large population, and that everyone in $\boldsymbol{v}$ belongs to one specific category $C_f \subset \boldsymbol{v},\ f = 1, \cdots, \rho$ with probability equal to

$$p_f := \mathbb{E}_{\boldsymbol{v}} \frac{|C_f|}{n},\ \ f = 1, \cdots, \rho. \tag{4}$$

Hence the set of $p_f,\ f = 1, \cdots, \rho$ describes the categorization-based population statistics, i.e., the statistics of the population from where $\boldsymbol{v}$ is randomly picked.

Without loss of generality it is assumed that the subject of interest belongs to the first category, i.e., that $v' \in C_1$. The fact that $v' \in C_1$ is also assumed to be known to the system - in other words, the system knows what it is looking for[6]. In this setting, pruning employs a categorization/estimation algorithm which, correctly or incorrectly, assigns each subject $v \in \boldsymbol{v}$ to a specific category

[5]We here assume that the pruning system is asked to leave at least one subject in $\mathcal{S}$, i.e., that $|\mathcal{S}| \geq 1$.

[6]We here clarify that asking for $v' \in C_1$ simply means that, prior to applying the presented results, the category of v' must be relabeled to C_1.

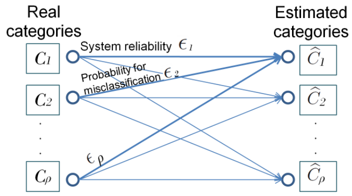

Figure 4. Confusion parameters $\{\epsilon_f\}$.

$\widehat{C}(v) \in [1, \rho]$. Errors may originate from say, algorithmic failures or reduced image quality. A subject $v \in \boldsymbol{v}$ is pruned out if and only if $\widehat{C}(v) \neq 1$, i.e., when it is estimated that v is not a member of C_1, whereas if $\widehat{C}(v) = 1$ then the subject v is not pruned out and is instead added into the selected set $\mathcal{S}$ of remaining candidates.

2.1.2 Error performance and confusion parameters

In concisely characterizing the error performance capabilities of the CV-based pruning system, we here adopt the simplifying common assumption that the confusion probability is defined by the categories, i.e., that for any subject $v \in C_f$, the probability that the categorization algorithm does not prune v, is a constant denoted as

$$\epsilon_f := P(\widehat{C}(v) = 1), \quad v \in C_f. \tag{5}$$

It becomes clear that ϵ_1 describes the system reliability (1 minus the probability of error), and also that for $f \geq 2$, ϵ_f describes the probability (cf. Fig. 4) that any member of C_f is misidentified to share the same characteristics as v', and is thus incorrectly not pruned out. We note that the adopted error measure, albeit an approximation, successfully reflects the fact that different categories may be easier to confuse than others[7]. Specifically having $\epsilon_f > \epsilon_{f'}$ means that subjects in category C_f can be more easily confused to belong to category C_1 of v', than subjects in category $C_{f'}$.

3. Statistical analysis using the method of types and information divergence

Let us consider a scenario where a search for a subject v' turned out to be extremely ineffective, and fell below the expectations, due to a very unfortunate matching of the subject with its surroundings $\boldsymbol{v}$. This unfortunate scenario motivates the natural question of how often will a system that was designed to achieve a certain average gain-reliability behavior, fall short of the expectations, providing an atypically small pruning gain and leaving its users with an atypically large and unmanageable $\mathcal{S}$. It consequently brings to the fore related questions such as for example, how will this probability be altered if we change the hardware and algorithmic resources of the system (change the ϵ_f and ρ), or change the setting in which the system operates (change the p_i).

[7]We note that the set $\{\epsilon_f\}$ is simply the first row of what is commonly known as a *confusion matrix*.

We proceed to analyze these issues and first recall that for a given authentication group $\boldsymbol{v}$, the categorization algorithm identifies set $\mathcal{S}$ of all unpruned subjects, defined as $\mathcal{S} = \{v \in \boldsymbol{v} \ : \ \widehat{C}(v) = 1\}$. We are here interested in the size of the search after pruning, specifically in the parameter

$$\tau := \frac{|\mathcal{S}|}{n/\rho}, \ 0 \leq \tau \leq \rho, \tag{6}$$

which represents[8] a relative deviation of $|\mathcal{S}|$ from a baseline n/ρ. It can be seen that the typical, i.e., common, value of τ is (cf. also Section 3.1)

$$\tau_0 := \mathbb{E}_{\boldsymbol{v}} \frac{|\mathcal{S}|}{n/\rho} = \rho \sum_{f=1}^{\rho} p_f \epsilon_f. \tag{7}$$

We are now interested in the entire tail behavior (not just the typical part of it), i.e., we are interested in understanding the probability of having an authentication group $\boldsymbol{v}$ that results in atypically unhelpful pruning ($\tau > \tau_0$), or atypically helpful pruning ($\tau < \tau_0$).

Towards this let

$$\alpha_{0,f}(\boldsymbol{v}) := \frac{|C_f|}{n/\rho}, \tag{8}$$

let $\boldsymbol{a}_0(\boldsymbol{v}) = \{\alpha_{0,f}(\boldsymbol{v})\}_{f=1}^{\rho}$ describe the *instantaneous* normalized distribution (histogram) of $\{|C_f|\}_{f=1}^{\rho}$ for the specific, randomly chosen and fixed authentication group $\boldsymbol{v}$, and let

$$\boldsymbol{p} := \{p_f\}_{f=1}^{\rho} = \{\mathbb{E}_{\boldsymbol{v}} \frac{|C_f|}{n}\}_{f=1}^{\rho}, \tag{9}$$

denote the *normalized statistical* population distribution of $\{|C_f|\}_{f=1}^{\rho}$.

Furthermore, for a given $\boldsymbol{v}$, let

$$\alpha_{1,f}(\boldsymbol{v}) := \frac{|C_f \cap \mathcal{S}|}{n/\rho}, \ 0 \leq \alpha_{1,f} \leq \rho, \tag{10}$$

let $\boldsymbol{\alpha}_1(\boldsymbol{v}) := \{a_{1,f}(\boldsymbol{v})\}_{f=1}^{\rho}$, and $\boldsymbol{\alpha}(\boldsymbol{v}) := \{\boldsymbol{\alpha}_0(\boldsymbol{v}), \boldsymbol{\alpha}_1(\boldsymbol{v})\}$, and let[9]

$$\mathcal{V}(\tau) := \Big\{0 \leq \alpha_{1,f} \leq \min(\tau, \alpha_{0,f}), \sum_{f=1}^{\rho} \alpha_{1,f} = \tau\Big\}, \tag{11}$$

denote the set of valid $\boldsymbol{\alpha}$ for a given τ, i.e., describe the set of all possible authentication groups and categorization errors that can result in $|\mathcal{S}| = \tau \frac{n}{\rho}$.

Given the information that $\boldsymbol{\alpha}_1$ has on $\boldsymbol{\alpha}_0$, given that τ is implied by $\boldsymbol{\alpha}_1$, and given that the algorithms here categorize

[8]Note the small change in notation compared to Section 1.4. This change is meant to make the derivations more concise.

[9]For simplicity of notation we will henceforth use $\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}, \alpha_{0,f}, \alpha_{1,f}$ and let the association to $\boldsymbol{v}$ be implied

a subject independently of other subjects, it can be seen that for any $\boldsymbol{\alpha} \in \mathcal{V}(\tau)$, it is the case that

$$\begin{aligned} P(\boldsymbol{\alpha}, \tau) &= P(\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1) = P(\boldsymbol{\alpha}_0) P(\boldsymbol{\alpha}_1 | \boldsymbol{\alpha}_0) \quad (12) \\ &= \prod_{f=1}^{\rho} P(\alpha_{0,f}) \prod_{f=1}^{\rho} P(\alpha_{1,f} | \alpha_{0,f}). \quad (13) \end{aligned}$$

The following lemma describes the asymptotic behavior of $P(\boldsymbol{\alpha}, \tau)$, for any $\boldsymbol{\alpha} \in \mathcal{V}(\tau)$. To clarify, the lemma describes the asymptotic rate of decay of the joint probability of an authentication group with histogram $\boldsymbol{\alpha}_0$ and an estimation/categorization process corresponding to $\boldsymbol{\alpha}_1$, given that the group and categorization process result in an unpruned set of size

$$|\mathcal{S}| = \tau \frac{n}{\rho} \quad (14)$$

for some $0 \leq \tau \leq \rho$. This behavior will be described below as a concise function of the binomial rate-function (cf.[10])

$$I_f(x) = \begin{cases} x \log(\frac{x}{\epsilon_f}) + (1-x) \log(\frac{1-x}{1-\epsilon_f}) & f \geq 2 \\ x \log(\frac{x}{1-\epsilon_1}) + (1-x) \log(\frac{1-x}{\epsilon_1}) & f = 1. \end{cases} \quad (15)$$

The lemma follows.

Lemma 1

$$-\lim_{n \to \infty} \frac{\log}{n/\rho} P(\boldsymbol{\alpha}, \tau) = \rho D(\boldsymbol{\alpha}_0 || \boldsymbol{p}) + \sum_{f=1}^{\rho} \alpha_{0,f} I_f(\frac{\alpha_{1,f}}{\alpha_{0,f}}),$$

where

$$D(\boldsymbol{\alpha}_0 || \boldsymbol{p}) = \sum_f \alpha_{0,f} \log \frac{\alpha_{0,f}}{p_f}$$

is the informational divergence between $\boldsymbol{\alpha}_0$ and $\boldsymbol{p}$ (cf. [10]).

The proof follows soon after. We now proceed with the main result, which averages the outcome in Lemma 1, over all possible authentication groups.

Theorem 1 *In SBS-based pruning, the size of the remaining set $|\mathcal{S}|$, satisfies the following:*

$$\begin{aligned} J(\tau) &:= -\lim_{n \to \infty} \frac{\log}{n/\rho} P(|\mathcal{S}| \approx \tau \frac{n}{\rho}) \\ &= \inf_{\boldsymbol{\alpha} \in \mathcal{V}} \rho \sum_{f=1}^{\rho} \alpha_{0,f} \log \frac{\alpha_{0,f}}{p_f} + \sum_{f=1}^{\rho} \alpha_{0,f} I_f(\frac{\alpha_{1,f}}{\alpha_{0,f}}). \quad (16) \end{aligned}$$

Furthermore we have the following.

Theorem 2 *The probability that after pruning, the search space is bigger (resp. smaller) than $\tau \frac{n}{\rho}$, is given for $\tau \geq \tau_0$ by*

$$-\lim_{n \to \infty} \frac{\log}{n/\rho} P(|\mathcal{S}| > \tau \frac{n}{\rho}) = J(\tau) \quad (17)$$

and for $\tau < \tau_0$

$$-\lim_{n \to \infty} \frac{\log}{n/\rho} P(|\mathcal{S}| < \tau \frac{n}{\rho}) = J(\tau). \quad (18)$$

The above describe how often we encounter authentication groups $\boldsymbol{v}$ and feature estimation behavior that jointly cause the gain to deviate, by a specific degree, from the common behavior described in (7), i.e., how often the pruning is atypically ineffective or atypically effective. We offer the intuition that the atypical behavior of the pruning gain is dominated by a small set of authentication groups, that minimize the expression in Theorem 1. Such minimization was presented in Fig. 2, and in examples that will follow after the proofs.

We now proceed with the proofs.

Proof of Lemma 1:

We first note that

$$P(\boldsymbol{\alpha}_0) \doteq e^{-nD(\boldsymbol{\alpha}_0/\rho || \boldsymbol{p})} = e^{-\frac{n}{\rho} D(\boldsymbol{\alpha}_0 || \rho \boldsymbol{p})} \quad (19)$$

where as previously stated $D(\boldsymbol{\alpha}_0 || \boldsymbol{p}) = \sum_f \alpha_{0,f} \log \frac{\alpha_{0,f}}{p_f}$ is the information divergence (also called the Kullback-Leibler distance) between $\boldsymbol{\alpha}_0$ and $\boldsymbol{p}$. We use $\doteq$ to denote exponential equality, i.e., we write $f(n) \doteq e^{-nd}$ to denote $\lim_{n \to \infty} \frac{\log f(n)}{n} = d$ and $\dot{\leq}$, $\dot{\geq}$ are similarly defined. In establishing $P(\boldsymbol{\alpha}_1 | \boldsymbol{\alpha}_0)$, we focus on a specific category f, and look to calculate

$$P\Big(|\mathcal{S} \cap C_f| = \frac{n}{\rho} \alpha_{1,f} \ \ | \ \ |C_f| = \frac{n}{\rho} \alpha_{0,f}\Big), \quad (20)$$

i.e., to calculate the probability that pruning introduces $\frac{n}{\rho} \alpha_{1,f}$ elements, from C_f to $\mathcal{S}$, given that there are $\frac{n}{\rho} \alpha_{0,f}$ elements of C_f. Towards this we note that there is a total of

$$|C_f| = \frac{n}{\rho} \alpha_{0,f} \quad (21)$$

possible elements in C_f which may be categorized, each with probability ϵ_f, to belong to C_1 by the categorization algorithm. The fraction of such elements that are asked to be categorized to belong to C_1, is defined by $\boldsymbol{\alpha}$ to be

$$x_f := \frac{|\mathcal{S} \cap C_f|}{|C_f|} = \frac{\frac{n}{\rho} \alpha_{1,f}}{|C_f|} = \frac{\alpha_{1,f}}{\alpha_{0,f}}, \quad (22)$$

an event which happens with probability

$$\begin{aligned} P(x_f) = P\Big(|\mathcal{S} \cap C_f| = \frac{n}{\rho} \alpha_{1,f} \ \ | \ \ |C_f| = \frac{n}{\rho} \alpha_{0,f}\Big) \\ \doteq e^{-|C_f| I_f(x_f)}, \quad (23) \end{aligned}$$

where in the above, $I_f(x_f) = x_f \log(\frac{x_f}{\epsilon_f}) + (1 - x_f) \log(\frac{1-x_f}{1-\epsilon_f})$ is the rate function of the binomial distribution with parameter ϵ_f (cf. [10]). Now given that

$$P(\boldsymbol{\alpha}_1 | \boldsymbol{\alpha}_0) = \prod_{f=1}^{\rho} P\Big(|\mathcal{S} \cap C_f| = \frac{n}{\rho} \alpha_{1,f} \ \ | \ \ |C_f| = \frac{n}{\rho} \alpha_{0,f}\Big) \quad (24)$$

we conclude that

$$-\lim_{n\to\infty}\frac{\log}{n/\rho}\log P(\boldsymbol{\alpha}_1|\boldsymbol{\alpha}_0)=\sum_f \alpha_{0,f}I_f(\frac{\alpha_{1,f}}{\alpha_{0,f}}). \quad (25)$$

Finally given that $P(\boldsymbol{\alpha},\tau)=P(\boldsymbol{\alpha}_0)P(\boldsymbol{\alpha}_1|\boldsymbol{\alpha}_0)$, we conclude that $-\lim_{n\to\infty}\frac{\log}{n/\rho}\log P(\boldsymbol{\alpha},\tau)=D(\boldsymbol{\alpha}_0||\rho\boldsymbol{p})+\sum_f \alpha_{0,f}I_f(\frac{\alpha_{1,f}}{\alpha_{0,f}})$.

□

Proof of Theorem 1: The proof is direct from the *method of types* (cf. [10]), which applies after noting that $|\mathcal{V}(\tau)|\leq n^{2\rho}\dot{\leq}e^{n\delta}\ \forall\delta>0$, and that $\sup_{\boldsymbol{\alpha}\in\mathcal{V}(\tau)}P(\boldsymbol{\alpha})\leq P(\tau)\leq|\mathcal{V}(\tau)|\sup_{\boldsymbol{\alpha}\in\mathcal{V}(\tau)}P(\boldsymbol{\alpha})$.

□

Proof of Theorem 2: The proof is direct by noting that for any $\delta>0$, then for $\tau\geq\tau_0$ we have

$$-\lim_{n\to\infty}\frac{\log}{n/\rho}P(|\mathcal{S}|>(\tau+\delta)\frac{n}{\rho})>-\lim_{n\to\infty}\frac{\log}{n/\rho}P(|\mathcal{S}|>\tau\frac{n}{\rho}), \quad (26)$$

and similarly for $\tau<\tau_0$ we have

$$-\lim_{n\to\infty}\frac{\log}{n/\rho}P(|\mathcal{S}|<(\tau-\delta)\frac{n}{\rho})>-\lim_{n\to\infty}\frac{\log}{n/\rho}P(|\mathcal{S}|<\tau\frac{n}{\rho}). \quad (27)$$

□

The following examples are meant to provide insight on the statistical behavior of pruning.

Example 2 (How often will the gain be very small?) *We recall the discussion in Section 1.4 where a pruning (surveillance) system can identify $\rho=3$ categories, operates with reliability $\epsilon_1=0.8$ over a population with statistics $p_1=0.4, p_2=0.25, p_3=0.35$ and has confusability parameters $\epsilon_2=0.2, \epsilon_3=0.3$. In the context of the above theorems we note that the result already shown in Fig. 2 applies by substituting τ with $\tau/\rho=\tau/3$. Consequently from Fig. 2 we recall the following. The size of the (after pruning) set $\mathcal{S}$ is typically 47.5% of the original size n. The probability that pruning removes less than $1-0.72=28\%$ of the original set, is approximately given by $e^{-\rho n}=e^{-3n}$ because, as Fig. 2 shows, $J(0.72)\approx 1$ (recall $\rho=3$). Similarly the same figure tells us that the probability that pruning removes less than $1-0.62=38\%$ of the original set, is approximately given by $e^{-\rho n/2}=e^{-3n/2}$ because $J(0.62)\approx 1/2$.*

In the following example we are interested in understanding the behavior of the search pruning in the case of rare authentication groups.

Example 3 (Which groups cause specific problems?) *Consider the case where a (soft biometrics based) pruning system has $\rho=2$ identifiable categories, population probabilities $\boldsymbol{p}=[p,\ 1-p]$, and confusion probabilities $\boldsymbol{\epsilon}=[1-\epsilon,\ \epsilon]$ (this means that the probability that the first category is confused for the second, is equal to making the reverse error). We want to understand what types of authentication groups will cause our pruning system to prune out only, for example, a fourth of the population ($|\mathcal{S}|\approx 3n/4$). The answer will turn out to be that the typical groups that cause such reduced pruning, have 43% of the subjects in the first category, and the rest in the other category.*

To see this we recall that (see Theorem 1) $|\mathcal{S}|\approx\tau\frac{n}{2}|\mathcal{S}|=3n/4$ which implies that $\tau=3/2$. For α denoting the fraction of the subjects (in $\boldsymbol{v}$) that belong in the first category, and after some algebra that we ignore here, it can be shown that $\alpha=\frac{3-\tau}{5-\tau}$ which yields $\alpha=3/7\approx 43\%$.

A further clarifying example focuses on the case of a statistically symmetric, i.e., maximally diverse population.

Example 4 (Male or female?) *Consider a city with 50% male and 50% female population ($\rho=2, p_1=p_2=0.5$). Let the confusion probabilities as before to be equal, in the sense that ($\boldsymbol{\epsilon}=[1-\epsilon,\ \epsilon]$). We are interested in the following questions. For a system that searches for a male (first category), how often will the system prune out only a third of the population (as opposed to the expected one half)? How often will we run across an authentication group with $\alpha\triangleq a_{0,1}=20\%$ males, and then have the system prune out only 45% of the overall size (as opposed to the expected 80%)? As it turns out, the first answer reveals a probability of about $e^{-n/4}$, and the second answer reveals a probability of about $e^{-n/5}$. For $n\approx 50$, the two probabilities are about five in a million and forty-five in a million respectively.*

To see this, first note that $\alpha_{1,1}=\alpha\tau$. Then we have that

$$I(\alpha,\alpha_{1,1},\tau):=\lim_{n\to\infty}\frac{\log}{n/2}P(|\mathcal{S}|=\frac{n}{2}\tau,\alpha,\alpha_{1,1}),$$

$$\inf_\delta I(\alpha,\alpha_{1,1},\tau)=I(\alpha,\tau,\alpha_{1,1}=\alpha\tau)=$$

$$2\alpha\log 2\alpha+2(1-\alpha)\log 2(1-\alpha)+\tau\log\tau+(2-\tau)\log(2-\tau).$$

To see the above, just calculate the derivative of I with respect to $\alpha_{1,1}$. For the behavior of τ we see that $I(\tau)=\inf_\alpha\inf_{\alpha_{1,1}}I(\alpha,\alpha_{1,1},\tau)=I(\alpha=p_1,\alpha_{1,1}=p_1\tau,\tau)$ $=\tau\log\tau+(2-\tau)\log(2-\tau)$, which can be seen by calculating the derivative of $\inf_\delta I(\alpha,\alpha_{1,1},\tau)$ with respect to α.

3.1. Typical behavior: average gain and goodput

We here provide expressions for the average pruning gain as well as the goodput which jointly considers gain and reliability. This is followed by several clarifying examples.

In terms of the average pruning gain, it is a simple exercise to see that this takes the form $\mathcal{G}:=\mathbb{E}_{\boldsymbol{v},\boldsymbol{w}}\mathcal{G}(\boldsymbol{v})=\left(\sum_{f=1}^{\rho}p_f\epsilon_f\right)^{-1}$, and similarly the average relative gain takes the form $\mathbb{E}_{\boldsymbol{v},\boldsymbol{w}}r(\boldsymbol{v})=\sum_{f=1}^{\rho}p_f(1-\epsilon_f)$. We recall that reliability is given by ϵ_1.

In combining the above average gain measure with reliability, we consider the (average) *goodput*, denoted as $\mathcal{U}$, which for the sake of simplicity is here offered a concise form of a weighted product between reliability and gain,

$$\mathcal{U}:=\epsilon_1^{\gamma_1}\mathcal{G}^{\gamma_2} \quad (28)$$

for some chosen positive γ_1, γ_2 that describe the importance paid to reliability and to pruning gain respectively.

We proceed with clarifying examples.

Example 5 (Average gain with error uniformity) *In the* uniform error setting *where the probability of erroneous categorization of subjects is assumed to be equal to ϵ for all categories , i.e., where $\epsilon_f = \epsilon = \frac{1-\epsilon_1}{\rho-1}$, $\forall f = 2, \cdots, \rho$, it is the case that*

$$\mathcal{G} = \left(p_1 + \epsilon - p_1\epsilon\rho\right)^{-1}. \tag{29}$$

This was already illustrated in Fig. 3. We quickly note that for $p_1 = 1/\rho$, the gain remains equal to $1/p_1$ irrespective of ϵ and irrespective of the rest of the population statistics $p_f,\ f \geq 2$.

Example 6 (Average gain with uniform error scaling) *Now consider the case where the uniform error increases with ρ as $\epsilon = \frac{\max(\rho-\beta,0)}{\rho}\lambda$, $\beta \geq 1$. Then for any set of population statistics, it is the case that*

$$\mathcal{G}(\lambda) = \left(p_1[1 + (\rho-\beta)\lambda] + \frac{\rho-\beta}{\rho}\lambda\right)^{-1}, \tag{30}$$

which approaches $\mathcal{G}(\lambda) = (p_1[1 + (\rho-\beta)\lambda] + \lambda)^{-1}$ as ρ increases. We briefly note that, as expected, in the regime of very high reliability ($\lambda \to 0$), and irrespective of $\{p_f\}_{f=2}^{\rho}$, the pruning gain approaches $\frac{1}{p_1}$. In the other extreme of low reliability ($\lambda \to 1$), the gain approaches ϵ_1^{-1}.

We proceed with an example on the average goodput.

Example 7 (Average goodput with error uniformity) *Under error uniformity where erroneous categorization happens with probability ϵ, and for $\gamma_1 = \gamma_2 = 1$, the goodput takes the form*

$$\mathcal{U}(\epsilon) = \frac{\epsilon + (1-\epsilon\rho)}{\epsilon + p_1(1-\epsilon\rho)}. \tag{31}$$

To offer insight we note that the goodput starts at a maximum of $\mathcal{U} = \frac{1}{p_1}$ for a near zero value of ϵ, and then decreases with a slope of

$$\frac{\delta\mathcal{U}}{\delta\epsilon} = \frac{p_1 - 1}{[\epsilon + p_1(1-\rho\epsilon)]^2}, \tag{32}$$

which as expected[10] is negative for all $p_1 < 1$. We here see that $\frac{\delta}{\delta p_1}\delta\frac{\mathcal{U}}{\delta\epsilon}|_{\epsilon\to 0} \to \frac{2-p_1}{p_1^3}$ which is positive and decreasing in p_1. Within the context of the example, the intuition that we can draw is that, for the same increase in ϵ[11], a search for a rare looking subject (p_1 small) can be much more sensitive, in terms of goodput, to outside perturbations (fluctuations in ϵ) than searches for more common looking individuals (p_1 large).

[10] We note that asking for $|\mathcal{S}| \geq 1$, implies that $\epsilon + p_1(1-\rho\epsilon) > \frac{1}{n}$ (cf.(29)) which guarantees that $\frac{\delta\mathcal{U}}{\delta\epsilon}$ is finite.

[11] An example of such a deterioration that causes an increase in ϵ, can be a reduction in the luminosity around the subjects.

4. Conclusions

The work provided statistical analysis of the gain and reliability in pruning the search over large data sets, where these sets are random and where there is a possibility that the pruning may entail errors. In this setting, pruning plays the role of pre-filtering, similar to techniques such as video indexing. The analysis may offer insight on better designing pre-filtering algorithms for different search settings.

References

[1] M. J. Swain and D. H. Ballard, "Color Indexing," *Inter. Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.

[2] S. Agarwal, A. Awan and D. Roth, "Learning to detect objects in images via a sparse, part-based representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1475–1490, 2004.

[3] E. Nowak, F. Jurie and B. Triggs "Sampling strategies for bag-of-features image classification," *in Proc. of ECCV,* 2006.

[4] A. K. Jain, S. C. Dass and K. Nandakumar, "Can soft biometric traits assist user recognition?," *in Proc. of SPIE,* 5404: 561–572, 2004.

[5] A. K. Jain, S. C. Dass and K. Nandakumar, "Soft biometric traits for personal recognition systems," *in Proc. of ICBA,* 2004.

[6] N. Kumar, P. N. Belhumeur and S. K. Nayar, "FaceTracer: a search engine for large collections of images with faces," *in Proc. of ECCV,* 2008.

[7] N. Kumar, A. C. Berg, P. N. Belhumeur and S. K. Nayar, "Attribute and simile classifiers for face verification," *in Proc. of IEEE ICCV,* 2009.

[8] G. Givens, J. R. Beveridge, B. A. Draper and D. Bolme, "A statistical assessment of subject factors in the pca recognition of human faces," *in Proc. of IEEE CVPR,* 2003.

[9] E. Newham, *The biometric report*, SJB Services, New York, 1995.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd edition, ISBN: 0-471-24195-4, Wiley, 2006.

[11] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Info. Theory*, vol. 49, no. 5, May 2003.

[12] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Info. Theory*, vol. 48, no. 8, Aug. 2002.

[13] A. Dantcheva and C. Velardo and A. D'Angelo and J.-L. Dugelay, "Bag of soft biometrics for person identification. New trends and challenges," *Multimedia Tools and Applications*, vol. 51, no. 2, pp. 739 - 777, 2011.

[14] A. Dantcheva, J.-L. Dugelay, and P. Elia, "Person recognition using a bag of facial soft biometrics (BoFSB),"*in Proc. of IEEE MMSP,* 2010.

[15] F. Escolano, P. Suau, B. Bonev, *Information Theory in Computer Vision and Pattern Recognition*, 2nd printing, Springer, ISBN: 978-1-84882-296-2, 2009.

[16] J. Coughlan and A. L. Yuille. "A Large Deviation Theory Analysis of Bayesian Tree Search," *In Math. Methods in Computer Vision. IMA Volumes in Mathematics and its Applications*, vol. 133, pp. 1-17. Springer-Verlag New York Inc., 2003.