DTU Library

Fingerprint Entropy and Identification Capacity Estimation Based on Pixel-level Generative Modelling

Yankov, Metodi Plamenov; Olsen, Martin A.; Stegmann, Mikkel Bille; Christensen, Søren Sk; Forchhammer, Søren

Link back to DTU Orbit

Fingerprint Entropy and Identification Capacity Estimation Based on Pixel-level Generative Modelling

Metodi P. Yankov, *Member, IEEE*, Martin A. Olsen, Mikkel B. Stegmann, Søren Sk. Christensen, and Søren Forchhammer, *Member, IEEE*

Abstract—A family of texture-based generative models for fingerprint images is proposed. The generative models are used to estimate upper bounds on the image entropy for systems with small sensor acquisition. The identification capacity of such systems is then estimated using the mutual information between different samples from the same finger. Similar to the generative model for entropy estimation, pixel-level model families are proposed for estimating similarity between fingerprint images with a given global affine transformation. These models are used for mutual information estimation, and are also adopted to compensate for local deformations between samples. Finally, it is shown that sensor sizes as small as 52x52 pixels are potentially sufficient to discriminate populations as large as the entire world population that ever lived, given that a complexity-unconstrained recognition algorithm is available which operates on the lowest possible pixel level.

Index Terms—Fingerprint recognition, biometric identification, biometric capacity, biometric entropy, generative modelling

I. Introduction

FINGERPRINT recognition is one of the more popular methods for biometric recognition due to its excellent performance combined with high security levels. The low cost of fingerprint sensors allows them to be integrated in ever greater portion of devices that require and/or can be used for recognition, such as mobile phones, personal computers and smart cards. In order to keep the cost down and allow higher integration, small sensors are attractive to vendors. However, such sensors pose a challenge for standard fingerprint recognition systems [1].

Fingerprints are historically and typically described by ridge endings and bifurcations, also known as minutiae. For full-print size sensors, minutiae-based fingerprint recognition is now well-established. In such cases, the individuality of fingerprints can be estimated from e.g. the probability of occurrence (PoO) of a minutiae configuration in the world [2], [3], extracted by statistical modelling. The number of possible configurations on the other hand can be related to the pattern rate in the pattern recognition field [4]. Since the

Metodi P. Yankov is with Fingerprint Cards A/S, 2730 Herlev, Denmark, and with the Department of Photonics Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark, e-mail: (metodi.yankov@fingerprints.com, meya@fotonik.dtu.dk).

Martin A. Olsen, Søren Sk. Christensen and Mikkel B. Stegmann are with Fingerprint Cards A/S, 2730 Herlev, Denmark.

Søren Forchhammer is with the Department of Photonics Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark.

Manuscript received XXX; revised XXX.

pattern rate does not capture intra-class variability (such as noise and skin warping and distortion during acquisition), the probability of random correspondence (PRC) was introduced to measure individuality [2], [5]. The PRC measures how likely it is to classify two random samples to the same finger, rather than the likelihood that a sample appears in the world. A good overview of the attempts at estimating fingerprint individuality is provided in [6], [5], most of which rely on said minutiae configurations. However, most of these methods cannot be tailored to small fingerprint samples due to insufficient number of minutiae present in the sample [6]. The individuality estimates with such methods is in small-sensor cases grossly underestimated and does not reflect the identification capabilities of algorithms, which are not constrained to minutiae matching. In some recent works, the security of small-sensor minutiae-based fingerprint recognition systems is even challenged based on similar arguments [7].

It is already known that a large amount of discriminatory information is present in the level 3 features, such as pores and ridge contours [6, Ch. 4]. Pore distributions can be exploited for improved PRC estimates [8] and also improved small-sensor performance [9]. To the best of our knowledge, the individuality of fingerprint samples when estimated on the lowest level of detail (the pixel level) is presently unknown. Such estimates are of great interest to fingerprint sensor design since they pose an ultimate limit on the sensor size for a given biometric performance.

In this paper, mutual information (MI) and entropy estimation for binarized fingerprint images is considered with a focus on small sensor acquisition systems. The entropy provides a lower bound on the length of a code, which can be used to uniquely describe the space of fingerprint images, acquired with that system. The entropy is thus one estimate of fingerprint individuality. Entropy can also be used as a quality metric for fingerprints [10]. Pixel-level entropy is thus of general interest also to practical fingerprint systems. On the other hand, MI is a more suited metric for identification capacity estimation since it takes into account the intra-finger image variation due to non-ideal acquisition conditions. The notion of biometric information has been used in a similar manner for face recognition when feature representation is applied [11]. To our knowledge, there are no published studies on the pixel-level MI for fingerprints.

This paper is organized as follows. In Section II, the basic foundations of entropy and MI estimation is given,

together with the system model under consideration. The entropy bounds rely on texture-based generative modelling, which is introduced in Section III. Models are then proposed in Section IV for MI and thereby fingerprint identification system capacity estimation. In Section V, some bounds and estimates of the entropy and MI are provided for different fingerprint sensor configurations and databases. In Section VI, the implications of the estimates are discussed, and Section VII concludes the work.

II. SYSTEM MODEL AND INFORMATION THEORY

The system under investigation is depicted in Fig. 1. During enrolment, the biometric source S (in this paper a fingerprint) is captured by an acquisition system, digitized and further processed (e.g. filtered, enhanced, binarized, etc.) into the image X. The acquisition system can be represented by a probability distribution $p_{X|S}$, which includes noise, imperfections, and all the above processing. Usually, feature extraction is performed in order to reduce the dimensionality of the recognition task, storage requirements and processing complexity of the system, as well as to increase the robustness to noise and intra-finger variations due to skin deformation. The features of interest may be further processed and stored in memory. The stored data related to one identity is that identity's template. During recognition, the biometric source is acquired by a potentially different system (although for simplicity, in this paper the acquisition system is assumed the same during enrolment and verification) into the image Y, its features are extracted and compared to the feature set in the memory. When running in identification mode, the recognition system classifies the probe to one of the templates in the memory, while in verification mode, the system produces a binary decision of type accept/reject.

Feature extraction is not a strict requirement. In fact, feature extraction can be viewed as processing, which inevitably reduces the information content [12, Ch. 2.8] of the template, and thus degrades the theoretical recognition performance bounds. In this paper, information content of fingerprint *image* data is extracted.

A. Notation

Variables are denoted with capital letters, e.g. X. Realizations of variable with small letters, e.g. x for scalars and $\mathbf{x}$ for vectors (matrices), $\mathbf{x}_k$ being the $k-$th element in a sequence of outcomes, and $\mathbf{x}(i,j)$ being the element on the $i-$th row and $j-$th column of the matrix. The probability of the outcome x is $p(x)$. A set of elements (pixels) around the (i,j)-th pixel is denoted as $\mathcal{C}$, is also referred to as *context*, and the elements of the matrix on those pixels with $\mathbf{x}(\mathcal{C})$. Other notations are defined on occurrence. Acronyms are defined on occurrence and are also given in the Appendinx, together with the main variables.

B. Entropy Estimation

The entropy $\mathcal{H}(X)$ of a random variable X is defined as the expectation of $-\log$ of its probability density function, or
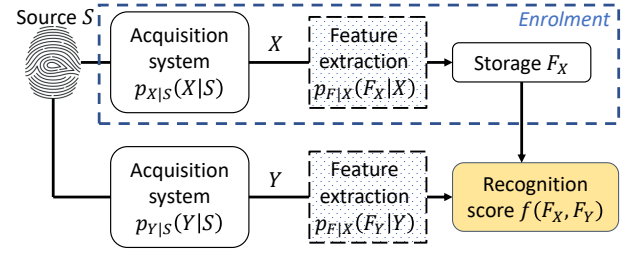

Fig. 1. Biometric recognition system model.

probability mass function (PMF) P_X for discrete variables. In the latter case

$$\mathcal{H}(X) = - \sum_{x \in supp(X)} p(x) \log_2 p(x), \qquad (1)$$

where $p(x) = P_X(X = x)$ for shorter notation and the entropy is measured in bits when the base of the logarithm is two. When X is the digitized biometric signal (as in Fig. 1), its entropy describes the pattern rate [4] for that biometric modality. In identification mode, the achievable rate $\mathcal{R}$ with which individuals can be discriminated with vanishing probability of error [13] (or equivalently, the maximum supported number of individuals $2^{\mathcal{R}}$) will be upper-bounded by the entropy as

$$\mathcal{R} \leq \mathcal{C} \leq \mathcal{H}(X), \qquad (2)$$

where $\mathcal{C}$ is the system capacity (defined below in (6)). The second inequality stems from the fact that the *acquisition and processing systems* are generally noisy. This difference is qualitatively similar to the difference between individuality estimated based on PRC and based on PoO of a fingerprint configuration, as described in the introduction. The first inequality stems from the fact that operating at capacity entails ideal *recognition system*. More details are given in the rest of this section. In this paper, X is an $m \times n$ dimensional digitized sample of the fingerprint S with a given resolution taking values in a discrete alphabet $\mathcal{X}^{m \cdot n}$ ($\mathcal{X}$ is the finite alphabet of each pixel, which can be e.g. binary or 8-bit depending on the quantization). When (1) cannot be solved in closed form, the entropy can be estimated from a long sequence of samples as

$$\mathcal{H}(X) = \lim_{K \to \infty} -\frac{1}{K} \sum_{k=1}^{K} \log_2 p(\mathbf{x}_k), \qquad (3)$$

where $\mathbf{x}_k$ is the $k-$th image in the set of K images, and the convergence to the true entropy is with probability 1 (cf. [14, Example 3] for application of (3) to telecommunications and references therein for applications to entropy in general). Furthermore, when the PMF is not available in closed form, the entropy can be upper-bounded by calculating (3) from the true samples $\mathbf{x}_1^K$, but replacing $P_X(X)$ with an auxiliary distribution $Q_X(X)$ [14]

$$\mathcal{H}(X) \leq \bar{\mathcal{H}}(X) = \lim_{K \to \infty} -\frac{1}{K} \sum_{k=1}^{K} \log_2 q(\mathbf{x}_k), \qquad (4)$$

where the closer Q_X is to P_X, the tighter the bound, with equality iff $Q_X(X = \mathbf{x}) = q(\mathbf{x}) = p(\mathbf{x})$ for all $\mathbf{x} \in supp(X)$. The auxiliary distribution Q_X can also be referred to as a *generative model*.

Examples of closed-form generative models for e.g. minutiae and pores are the above-mentioned [2] and [8], respectively, or the minutia model obtained with histograms [3]. A generative model for the ridge-flow can also be found in [15]. In this paper, the focus is on pixel-level modelling, which as argued below, improves the system capacity. Another example of generative models for biometrics can be found in [16] for the IrisCode, where a 1-st order Markov chain is used as an auxiliary generative model and for estimation of the IrisCode entropy. When the auxiliary model is simple, its entropy can be estimated in closed form from (1) (as also done in [16] and [3]). When the model is more sophisticated (as in the rest of the paper), sequence based entropy estimation should be performed (e.g. from (3)).

When the limit in (3) is approached with samples from the same finger, the entropy describes the entropy of that finger and is estimated based on the average PoO of its samples. Such representation is useful in terms of estimating uniqueness of a given finger. On the other hand, when the limit is approached with samples from infinitely many fingers S, the entropy represents the average entropy for that biometric modality. The entropy can also be estimated based on a single image, in which case it is just equivalent to the PoO of that sample.

C. Entropy of images

A particularly popular generative model for images is the causal pixel predictive model (PPM) [17], which attempts simplifying the joint distribution $q(\mathbf{x})$ into

$$
\begin{aligned}
q(\mathbf{x}) &= \prod_{(i,j)\in\{1:m,1:n\}} q(\mathbf{x}(i,j)|\mathbf{x}(\mathcal{C})) \\
&\triangleq \prod_{(i,j)\in\{1:m,1:n\}} q(\mathbf{x}(i,j)|\mathbf{x}(\hat{\mathcal{C}}(i,j))),
\end{aligned} \quad (5)
$$

where $\mathbf{x}(i,j)$ is the pixel at the $i-$th row and $j-$th column, $\mathcal{C}$ is the set of previously traversed pixels (also referred to as *past*), $\hat{\mathcal{C}}(i,j)$ is a small subset around the pixel of interest (PoI) (i,j), also known as *context*, and $\mathbf{x}(\hat{\mathcal{C}}(i,j))$ is the set of pixel values in the context. The first equality is simply the product rule for probabilities, followed by dimensionality truncation. In the case of $\hat{\mathcal{C}} \equiv \mathcal{C}$, $q(\mathbf{x}) = p(\mathbf{x})$. The PPM we apply in this paper is exemplified in Fig. 2 for a 4D context and a 12D context, for which (i,j) are traversed in a row-wise fashion, with the context only containing pixels in the past. Notice that other configurations of the context are possible, and this particular is chosen for consistency and easier implementation of context with variable radius. Since $q(\mathbf{x})$ is a valid distribution, the inequality in (4) still holds.

The entropy of images also describes their lossless compressability in that it lower-bounds the length of the code, which can be used to compress an image (or source in general) [12, Ch. 5.4].

D. Mutual Information Estimation

In [13], the system identification capacity from (2) is shown to be

$$
\mathcal{C} = \mathcal{I}(X;Y) = \mathcal{H}(X) - \mathcal{H}(X|Y), \quad (6)
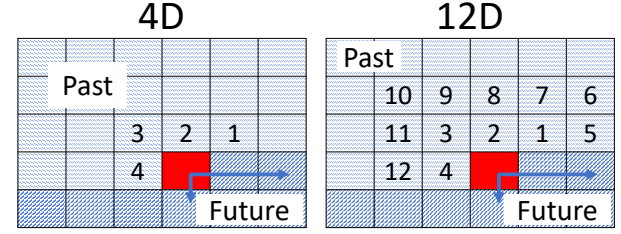$$

Fig. 2. PPM and image traversing illustration.

where $\mathcal{H}(X|Y)$ is the conditional entropy of X once Y is observed, and $\mathcal{I}(X;Y)$ is the MI between X and Y. In this context, the MI describes how much of the information in X is also present in Y. The conditional entropy $\mathcal{H}(X|Y)$ can be viewed as a measure of similarity between images, whereas $\mathcal{H}(X)$ represents the total amount of information (including noise), present in X. As mentioned, when estimated for a single image, $\mathcal{H}(X)$ is directly related to the image uniqueness through its PoO. Typically, recognition algorithms rely solely on similarity, thus implicitly estimating $\mathcal{H}(X|Y)$ and neglecting $\mathcal{H}(X)$. As seen from (6), theoretically both measures are required for approaching the system capacity, and their difference should be maximized for maximizing that capacity. As $\mathcal{H}(X|Y)$ is non-negative, the capacity (6) is ultimately bounded by $\mathcal{H}(X)$. The MI representation of capacity can be understood with the following simple examples:f

1) *Ideal acquisition system and fixed acquisition conditions.* In this case, the variability between samples of the same finger is zero, the images X, Y and S are identical, $\mathcal{H}(X|Y) = 0$ and the entire information content of S can be used for biometric recognition, resulting in a high capacity.
2) *Completely noisy system.* In this case, X and Y are independent, $\mathcal{H}(X|Y) = \mathcal{H}(X)$ and $\mathcal{R} = \mathcal{C} = 0$.
3) *Extremely robust system.* In this case, zero (or small) variability can be achieved at the expense of sacrificing (e.g. by quantizing) most of the details of S. The result is identical X and Y and thereby small $\mathcal{H}(X|Y)$. However, in such cases the entropy $\mathcal{H}(X)$, and thus capacity, is also small.

Minutiae based recognition systems can be classified to a moderate example 3), since details in the image are sacrificed (not utilized), in order to exploit the robustness and repeatability of the minutia configuration. Another example of a robust feature is the orientation map (OMAP). The MI between OMAPs of two images can be used for relative transformation estimation [18], but a system, relying solely on OMAP for recognition would be penalized by the low entropy. Feature-based biometric information was estimated for face samples [11], where it is also used to estimate system capacity. As mentioned, feature-based MIs do not fully exploit the information content available at the pixel-level, but can be utilized to improve the pixel-level modelling, as will be seen below.

Similar to $\mathcal{H}(X)$, $\mathcal{H}(X|Y)$ can be estimated and bounded

as

$$\mathcal{H}(X|Y) = \lim_{K\to\infty} -\frac{1}{K}\sum_{k=1}^{K}\log_2 p(\mathbf{x}_k|\mathbf{y}_k) \qquad (7)$$

$$\leq \lim_{K\to\infty} -\frac{1}{K}\sum_{k=1}^{K}\log_2 q(\mathbf{x}_k|\mathbf{y}_k), \qquad (8)$$

where $p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{y})} = \frac{\sum_{\mathbf{s}\in\mathcal{S}} p(\mathbf{s})p(\mathbf{x}|\mathbf{s})p(\mathbf{y}|\mathbf{s})}{\sum_{\mathbf{s}\in\mathcal{S}} p(\mathbf{s})p(\mathbf{y}|\mathbf{s})}$ [13] and $Q_{X|Y}(X=\mathbf{x}_k|Y=\mathbf{y}_k) = q(\mathbf{x}_k|\mathbf{y}_k)$ is any valid distribution. In this paper, auxiliary distributions of the form

$$q(\mathbf{x}|\mathbf{y}) = \prod_{(i,j)\in\{1:m,1:n\}} q(\mathbf{x}(i,j)|\mathbf{x}(\mathcal{C}),\mathbf{y})$$

$$\triangleq \prod_{(i,j)\in\{1:m,1:n\}} q(\mathbf{x}(i,j)|\mathbf{x}(\hat{\mathcal{C}}(i,j)),\mathbf{y}(\dot{\mathcal{C}}(i,j))), \quad (9)$$

are adopted, where the context $\dot{\mathcal{C}}(i,j)$ is not necessarily identical to $\hat{\mathcal{C}}(i,j)$ and can include any subset of $\mathbf{y}$, including future pixels.

The auxiliary distribution $Q_{X|Y}$ will be referred to as a *channel model*, highlighting the similarities between biometric recognition and communications technology. In the latter case, a source X is communicated through a channel $p_{Y|X}$, and the maximum error-free transmission rate is given by $\mathcal{I}(X;Y)$. However, p_X and $p_{X|Y}$ are typically available, and $\mathcal{I}(X;Y)$ for standard channels and sources is thus straight forward to calculate. In biometrics, both distributions are typically unknown, and must be thus replaced by their auxiliary versions. In this paper, the term *channel* is abstract (as no actual communication channel exists) and serves the purpose of representing the relation between the images X and Y.

An example is given in Fig. 3 for a 4D generative model (5) and a 13D channel model (9) contexts. For simplicity, the radii of both models in both images are assumed identical, which is also not a requirement.

In telecommunications, achieving the capacity entails using an optimal and infinitely long error-correcting code. Similarly, as mentioned above, the entropy only provides a bound on the length of the source code, which can be used to compress the signal. The design of such error-correcting and source codes is by far a non-trivial task. In the case of biometrics in general, and fingerprints in particular, error-correcting and source code design can be related to the design of optimal feature vectors. Such feature vectors allow to extract the useful information from the image in an efficient manner (source code), and at the same time are encoded for maximum separability of the classes (error-correcting code). Such design falls outside the scope of the present paper.

III. GENERATIVE MODELS

The first class of models we apply are the look-up table (LUT) models, for which the PMF $q(\mathbf{x}(i,j)|X(\hat{\mathcal{C}}) = \mathbf{x}(\hat{\mathcal{C}}))$ is defined for each possible realization of the context. This model is trained by traversing training images and building histograms for each realization. The complexity of this model grows exponentially with the size of the context (since the number of realizations is $|supp(X(\hat{\mathcal{C}}))| = |\mathcal{X}|^{|\hat{\mathcal{C}}|}$). Such
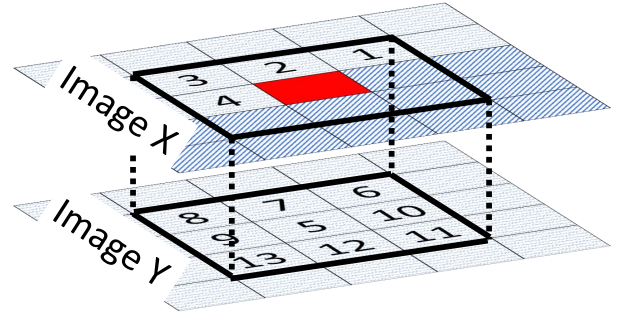

Fig. 3. PPM with probe and reference for MI estimation with a 4D generative model and the respective $(3\cdot4+1)$D channel model.

models are therefore typically limited to binary alphabets and small contexts (e.g. 10 pixels in the binary image compression standard JBIG [19] or up to 16 pixels in JBIG2 [20]).

A. Closed-form expression models

In order to extend the context and the pixel alphabet, closed-form expression models are required. In this work, two such models are studied: a logistic regression (LR) linear model, and a neural network (NN), which operate in classification mode and are able to produce log-likelihoods of each class [21]. The output classes in this case represent $\mathcal{X}$. Both models are trained by traversing training images similar to the LUT model. The training data consist of input/output pairs of points, where the input points are the $|\hat{\mathcal{C}}|-$dimensional context of each PoI in each training image, and the output is each said PoI. The LR model is trained using the LIBLINEAR optimization libraries [22]. Single layer NN configuration is studied in this work with a hidden layer size optimized to 50 units. The stochastic gradient-based algorithm [21, Ch. 5] is used for NN training.

B. Including orientation map

The generative models can be improved by including the local ridge-flow OMAP in the context. This is equivalent to estimating the MI $\mathcal{I}(X;Y|O)$ instead, where O describes the OMAP. It is noted that as mentioned above, the OMAP is a feature itself and carries discriminatory information, which can be used to improve the matching performance. Conditioning on the OMAP essentially provides estimates on the performance of algorithms, which assume the OMAP is known prior to matching, e.g. from estimation, and do not exploit it explicitly for decision making. In this paper, the OMAP is estimated with standard gradient-based methods [6, Ch. 3], and is assumed available when estimating the entropy (with alphabet spanning the integer range 1-180 degrees). It is noted that generative models including the OMAP have double dimensionality. We also note that the OMAP of all pixels in the context is not as necessary as texture information, and can be replaced by e.g. the average OMAP of the context. However, for consistency, the local OMAP of all context pixels is used in this work.

IV. CHANNEL MODELS

The channel models as described in Section II exhibit texture dimensionality of three times the dimensionality of the generative model plus one pixel (see Fig. 3). Even for small radii, LUT models are thus infeasible. For channel models, only the NN is applied in this paper, which also includes the orientation map of image X. It is assumed that the orientation of image Y is identical, or at least sufficiently similar to that of image X to not bring additional information for genuine matches, and is thus not included in the model in order to reduce complexity.

Training and testing images X and Y are generated by running a commercial recognition algorithm for genuine matches as described below. The MI estimation process is summarized in Algorithm 1. The recognition algorithm determines the global transformation T between images and transforms (translates and rotates) the image $Y = T(Y)$ to match the image X. A mask image M is then constructed for the overlapping area ($\mathbf{m}(i,j) = 1$ if the images overlap at that pixel), and the pixels in those areas are used for training of the NN and for conditional entropy and MI estimation.

Examples of this process are given in Section V. An example of such an NN used as channel model is given in Fig. 4. The generative model is identical, but of lower dimensionality as it does not include the inputs $\mathbf{y}(\dot{\mathcal{C}}(i,j))$.

V. RESULTS

A. Dataset description

An overview of the datasets used to study the identification capacity is given in Table I. The FVC databases are comprised of 100 fingers, eight impressions each [6]. The MCYT databases [23], [24] contain 1000 fingers, 12 impressions each, but for consistency, only the first 100 fingers are used. As argued below, they are sufficient to see convergence in the entropy and MI estimates. Each image from the datasets is cropped at a random location to a desired sensor size, ensuring that the captured area does not contain empty spaces (condition was 99% active area). Crop sizes of 80x80, 120x120 and 160x160 pixels are studied in this work. The training dataset is comprised of the first 50 fingers, and entropy estimation is performed on the rest. Smaller crops are not studied since this leads to insufficient training and testing dataset size.

A synthetic database generated with the SFinGe model is also used [6, Ch. 6], [25]. Two types of images are generated: 1) ridge-line pattern; 2) noisy and distorted realistic images. It is noted that this database assumes that the only discriminatory information in fingerprints is minutiae (pores are not generated), and disregards texture information which as mentioned is in fact highly discriminatory. The noisy realistic images are generated with an average image quality distribution as defined by SFinGe. Similar to the public data base, small-sensor images are produced by cropping full-size prints. The training set in this case similarly comprises 400 images, with 400 testing images.

For MI estimation, one image per finger is used as template (image X) and the rest (images Y) are matched against it with a commercial algorithm. Training and testing images are
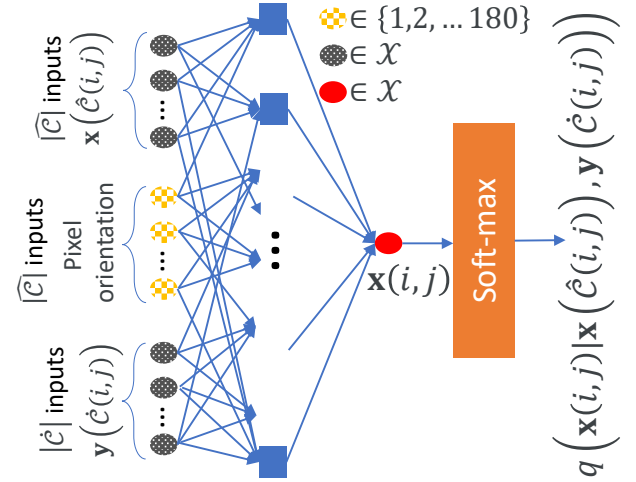

Fig. 4. An example of an NN channel model for the auxiliary distribution $q(\mathbf{x}|\mathbf{y})$.

TABLE I
SUMMARY OF THE DATABASES USED IN THIS WORK.

	Acquisition technology	Array size, pixels	Resolution, dpi
FVC-2002-DB1-A	Optical	388x374	500
FVC-2002-DB2-A	Optical	296x560	569
FVC-2002-DB3-A	Capacitive	300x300	500
FVC-2004-DB1-A	Optical	640x480	500
FVC-2004-DB2-A	Optical	328x364	500
FVC-2004-DB3-A	Thermal	300x480	512
MCYT optical	Optical	256x400	500
MCYT capacitive	Capacitive	300x300	500
SFinGe	Synthetic	416x560	500

then generated as described in Section IV. It is noted that the random cropping sometimes results in insufficient overlap for our algorithm to correctly match the probe to the reference. In order to mitigate this, only images of size 160x160 are used for conditional entropy and MI estimation. However, the training and testing datasets in this case are still somewhat smaller than for the entropy estimation, since the entire images are not used, but only the areas, where the image X overlaps with image Y.

In all cases, the training size is more than $2.5 \cdot 10^6$ pixels, and testing size is more than $1.5 \cdot 10^6$ pixels.

B. Entropy estimates for small sensors

In Fig. 5, the entropy estimates for the FVC-2002-DB3-A database are shown as a function of the context size for the 80x80 cropped samples. The studied sizes of 4, 12, 24, 40, 84 and 112 correspond to context radii of 1, 2, 3, 4, 5 and 6 pixels (see Fig. 2). The LUT model is limited by its complexity to 12 pixels, whereas the closed-form expression models (from Section III-A) were possible to simulate on a standard PC for up to 112 pixels, including the double dimensionality OMAP case. In order to reduce the number of possible context configurations for the LUT case, the rather small images of size 80x80 are first rotated s.t. the average OMAP of the entire image is zero. We found that this significantly tightens the entropy bounds for the LUT model. For the small-context size, the LUT is slightly better than the closed form models,

---

Algorithm 1 Algorithm for estimation of MI between two fingerprint images.

Require: Images X, Y
1: Find transform and mask $[T, \mathbf{m}] = match(X, Y)$ ▷ 'match' is the recognition algorithm of choice
2: Transform $Y = T(Y)$
3: Extract OMAP **o**   ▷ Orientation of image X
4: $\mathcal{H}(X) = 0, \mathcal{H}(X|Y) = 0, \mathcal{I}(X;Y) = 0, N_{pixels} = 0$
5: **for** $i = 1 : N_{rows}$ **do**   ▷ Traverse rows
6:      **for** $j = 1 : N_{cols}$ **do**   ▷ Traverse columns
7:         **if** $\mathbf{m}(i, j)$ **then**
8:            Extract contexts (NN inputs) $\mathbf{x}(\hat{\mathcal{C}}(i,j))$, $\mathbf{o}(\hat{\mathcal{C}}(i,j))$ and $\mathbf{y}(\dot{\mathcal{C}}(i,j))$
9:            **if** Training mode **then**
10:              Store NN inputs and NN output $\mathbf{x}(i, j)$ for NN training
11:            **else if** Estimation mode **then**
12:              Estimate PoO $q(\mathbf{x}(i,j)|\mathbf{x}(\hat{\mathcal{C}}(i,j)), \mathbf{o}(\hat{\mathcal{C}}(i,j)))$ ▷ Generative model
13:              Estimate PoO $q(\mathbf{x}(i,j)|\mathbf{x}(\hat{\mathcal{C}}(i,j)), \mathbf{y}(\dot{\mathcal{C}}(i,j)), \mathbf{o}(\hat{\mathcal{C}}(i,j)))$ ▷ Channel model, Fig. 4
14:              $\mathcal{H}(X) = \mathcal{H}(X) - \log_2 q(\mathbf{x}(i,j)|\mathbf{x}(\hat{\mathcal{C}}(i,j)))$
15:              $\mathcal{H}(X|Y) = \mathcal{H}(X|Y) - \log_2 q(\mathbf{x}(i,j)|\mathbf{x}(\hat{\mathcal{C}}(i,j)), \mathbf{y}(\dot{\mathcal{C}}(i,j)))$
16:              $N_{pixels} = N_{pixels} + 1$
17:            **end if**
18:         **end if**
19:      **end for**
20: **end for**
21: $\mathcal{H}(X) = \mathcal{H}(X)/N_{pixels}, \mathcal{H}(X|Y) = \mathcal{H}(X|Y)/N_{pixels}$ ▷ Normalize
22: $\mathcal{I}(X;Y) = \mathcal{H}(X) - \mathcal{H}(X|Y)$

---

which can be expected as the histogram can be made very accurate with the amount of training data available (making Q_X a better approximation of P_X). When the dimensionality is increased, both the NN and LR models improve the tightness of their bounds, reaching a saturation at around 60 pixel contexts. When the OMAP is included in the model, the LR-based bound does not improve, while the NN is able to tap into this fingerprint domain property and model the images better. An optimal value of 40 pixels is found, leading to model dimensionality of 80. The corresponding diameter of the context is 9 pixels, which at 500 dpi is sufficient to capture a ridge-valley period [6, Sec. 6.4.3]. The NN thus can conceivably model the ridge-valley frequency as a macro-feature. When the context size is increased, the NN-based entropy bound becomes unstable. This can be attributed to estimation errors and overfitting of the very complex model to the limited amount of training data. The lowest upper bound we could find for this dataset is ≈ 0.3552 bits/pixel. For a sensor size of 80x80 this means an average entropy of 2273 bits/sample.

For the rest of the paper, the context size in the generative model is set to 40 pixels (radius of four pixels), the orientation map is included and the NN model is used.

In Fig. 6, the histogram of the finger entropy estimates with bin size of 0.05 bits is studied (the set of K images in (3) includes only the eight images of the same finger). As mentioned, the inter-finger variation can be interpreted as variable degree of uniqueness for each finger, i.e., some fingers will experience higher false acceptance rate when matching (the "lambs" in the Doddington zoo context [26]) due to their smaller entropy (related to their higher PoO). In the same figure, the effect of sensor size is also studied. As seen, the
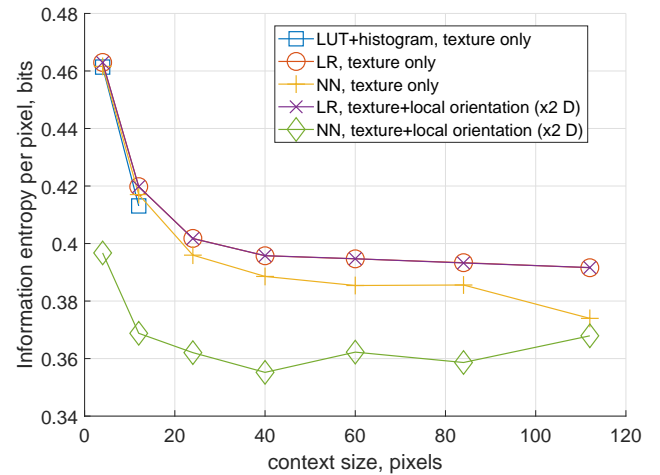

Fig. 5. Entropy estimates vs. context size for the studied generative models.

estimates (on average) are very similar regardless of the sensor size. This suggests that individuality, and thus recognition performance will be a function largely of the reference image coverage area, and is independent of the sensor size at least down to the smallest studied crop size of 80x80. Thus, if a probe image has an overlap of sufficient size with a reference image (governed by the entropy per pixel), the biometric performance does not suffer from the small sensor.

Finally in this section, the border effect is discussed. In the above, in order to initialize the traversal, the top and left borders, with width defined by the radius of the context, are assumed known. In practice, the entropy and MI per pixel reported in this paper will have to be adjusted, e.g. by rescaling the per-pixel results or by assuming a slightly larger sensor.
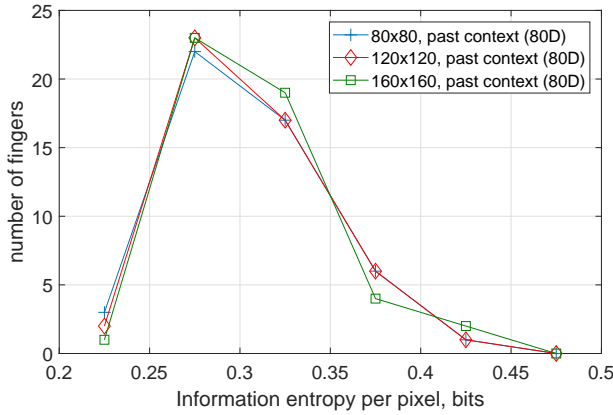
Fig. 6. Histogram of the entropy estimates per finger for sensors of different size.

For consistency, in this paper only the inner parts of the images are used for entropy and MI estimation and the border effect described is neglected.

C. Mutual Information Estimates

In Fig. 7, an example is given for the MI estimation between two images. Once the global transformation is found, the images are traversed as in Fig. 2, and the conditional entropy is estimated for the overlapping areas following the context illustration in Fig. 3. The probability map $p(X|Y)$ is generally brighter than the map $p(X)$, which is a result of the similarity between the image areas. Observe that the mutual information is contained in pixels, which carry a lot of self-information (low PoO $p(X)$) and at the same time are similar in both images (high PoO $p(X|Y)$). On the other hand, extremely noisy areas (having low PoO $p(X)$), will be very different in image Y (resulting in low PoO $p(X|Y)$ as well), and will thus not contribute to the MI, since the terms cancel in Eq. (6).

In Table II, the average entropy and MI per pixel and per area are given for all studied databases. The *entropy* estimates reported for FVC-2002-DB3-A are slightly lower than the ones reported in Fig. 5, since the entropy in the former case is estimated only for the matched areas. The recognition algorithm that was used favours less-noisy areas, and suffers performance penalty in areas, which are more noisy, and thus higher in entropy content. Such areas are thus used less often for the estimate in this section, since they appear less often in the dataset with matched images.

The ridge-line synthetic SFinGe images exhibit significantly smaller entropy than that of captured and realistic images. There are several reasons for this discrepancy:

1) There are noise and distortions in real images, which do not carry discriminatory information, however, their entropy is added to the estimates.
2) As mentioned, ridge-lines and minutiae do not carry the complete biometric information of a fingerprint. Some discriminatory information is thus not present in the ridgeline dataset.
3) In a less noisy environment, the generative models can be trained much more accurately, leading to tighter entropy bound.

TABLE II
COMPARISON OF ENTROPY UPPER BOUNDS AND MI ESTIMATES FOR THE STUDIED DATABASES (DESCRIBED IN THE TEXT).

	Entropy per pixel	Entropy per mm^2	MI per pixel	MI per mm^2
FVC-2002-DB1-A	0.299	119.6	0.0188	7.52
FVC-2002-DB2-A	0.258	124.9	0.0158	7.64
FVC-2002-DB3-A	0.318	127.2	0.0196	7.84
FVC-2004-DB1-A	0.250	100.0	0.0233	9.32
FVC-2004-DB2-A	0.269	107.6	0.0156	6.24
FVC-2004-DB3-A	0.252	102.3	0.0135	5.48
MCYT optical	0.341	136.6	0.0374	4.96
MCYT capacitive	0.345	138.2	0.035	14.01
SFinGe, ridge-lines	0.193	77.2	0.057	22.80
SFinGe, realistic	0.315	126.0	0.057	22.80

This being said, there exist variations across databases, which are attributed mostly to different quality, acquisition technology and dataset collection environment and purpose. Optical sensors generally exhibit lower entropy bounds, as expected from the higher-quality, less-noisy acquisition. In the case of FVC-2002-DB2-A, the lower entropy *per pixel* is also attributed to the higher resolution. The difference in the entropy *per area* estimates for FVC-2002-DB1-A and FVC-2002-DB2-A is $\approx 4\%$. The MCYT databases exhibit significantly higher MI, which is attributed to the higher image quality. Even-though the sensor used to collect the MCYT optical database is the same model as that of FVC-2004-DB2-A, the latter database was designed to be as realistic as possible, while the former contains both random and highly-controlled captures [23], resulting in high MI. Furthermore, the sensor was not cleaned during the FVC collections, which is also clearly visible upon visual inspection of the images, resulting in poorer overall quality. Similar arguments can be made for the difference in MI between the MCYT capacitive database and FVC-2002-DB3-A, which employ the same sensor model for acquisition. [1]

The MIs of the SFinGe ridgeline and noisy samples are remarkably close, regardless of their significantly different entropies. This highlights the validity of the NN as a channel model, as it is able to successfully extract the interesting portion of information from the noisy samples and disregards the information content of the noise part. The higher MI for synthetic images has two possible explanations:

1) The particular noise configuration generated by SFinGe for this paper is in fact more optimistic than what true sensors provide.
2) The underlying fingerprint distribution defined by SFinGe is significantly different than that of true datasets, and results in unrealistically high capacity expressed by MI.

Finally in this subsection, the convergence of the estimates is studied. In Fig. 8, the estimates are given as a function of the number of pixels, used in the estimation for the FVC-2002-DB1 database. The vertical lines indicate bulks of ten images (unevenly distributed due to different size of the overlap area

---

[1]Typical images from FVC can be found on http://bias.csr.unibo.it/fvc2004/databases.asp and http://bias.csr.unibo.it/fvc2002/databases.asp, and typical MCYT images in [23], and are omitted here for space.
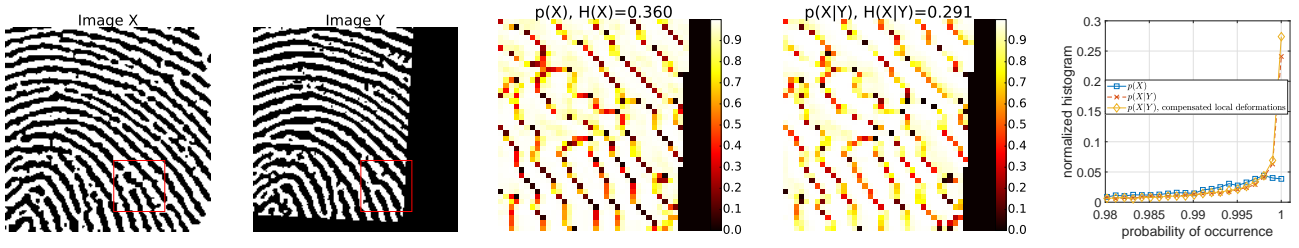
Fig. 7. Example images used for estimation of the entropies of the overlapping area $\mathcal{H}(X) = 0.360$, $\mathcal{H}(X|Y) = 0.291$ and the MI of the overlapping areas $\mathcal{I}(X;Y) = \mathcal{H}(X) - \mathcal{H}(X|Y) = 0.069$ bits/pixel. An example of the estimated pixel PoO is given for a small patch area, based on the generative and channel models. **Right-most figure:** A histogram of the PoO for the generative model and the channel model, including the case when local deformations are taken into account, as in Section V-D. In the latter case, more pixels are explained by image Y, resulting in even lower $\mathcal{H}(X|Y)$ and higher $\mathcal{I}(X;Y)$ (0.259 and 0.101 bits/pixel, respectively, for the example images X and Y). Images taken from FVC-2002 DB3.

for each match). We see convergence after $\approx 1.2 \cdot 10^6$ pixels, which is achieved with ≈ 90 images. As mentioned, more than $1.5 \cdot 10^6$ pixels are used for estimation in each case, deeming the estimates converged.

D. Local deformations compensation

Real-life captures are usually subjected to local deformations, making the globally found transformation between images sub-optimal locally. In order to compensate for this effect, a non-stationary context is applied by searching the local area of pixel (i,j) in image Y for a better match to the context $\mathbf{x}(\hat{\mathcal{C}}(i,j))$. This is achieved by "jiggling" image Y around pixel (i,j) in order to find a better match to image X at that neighbourhood. The search is done with a radius R, and the context of image Y is centred around the pixel (k,l)

$$(k,l) = \arg \min_{k \in i-R:i+R, l \in j-R:j+R} d_H(\mathbf{x}(\hat{\mathcal{C}}(i,j)), \mathbf{y}(\hat{\mathcal{C}}(k,l))), \tag{10}$$

where $d_H(\cdot,\cdot)$ is the Hamming distance between two vectors. The context $\mathbf{y}(\dot{\mathcal{C}}(k,l))$ then replaces $\mathbf{y}(\dot{\mathcal{C}}(i,j))$ in Eq. (9) and changes from pixel to pixel. It should be noted that stationarity is one of the requirements for the convergence in (7), and the resulting estimate of $\mathcal{H}(X|Y)$ is thus no longer a bound. However, it does provide an improved estimate of the achievable rate in practice if the local deformations are taken into account. Notice also that past contexts $\hat{\mathcal{C}}$ are used for estimating (k,l) in (10), whereas past+future contexts $\dot{\mathcal{C}}$ are used for conditional entropy estimation in (9).

The size of the area used for estimating the minimum distance in (10) has a major effect on the validity of the MI estimation. When the area is small w.r.t. e.g. the ridge frequency or the generative model context size, increasing the search radius beyond what can be considered 'local' still brings benefits in terms of finding an area with reduced Hamming distance to the context in image X. This effect is exemplified in Fig. 9, where the entropy, conditional entropy and MI are shown for FVC-2002-DB3-A as a function of the search radius for different size of the match area $\hat{\mathcal{C}}$ from (10). The match area should not be confused with the context size of the generative model, which is kept at 40 pixels, and with the context size of the channel model, kept at $3 \cdot 40 + 1$ pixels (see Fig. 3). Small match area results in increasing MI with the search radius, which however cannot be considered local deformation compensation, as the match area falls outside the
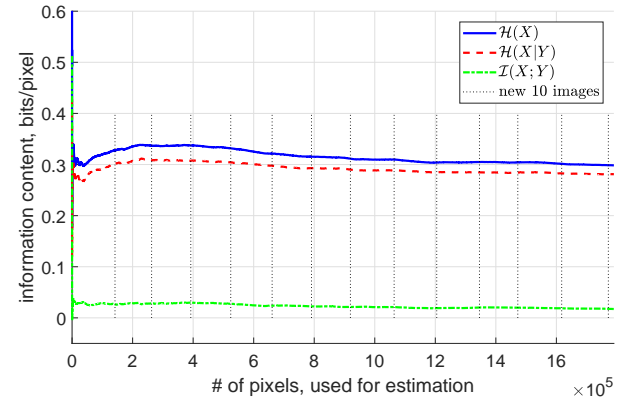

Fig. 8. Convergence properties of the entropy and MI estimation for FVC-2002-DB1. Convergence is achieved after $\approx 1.2 \cdot 10^6$ pixels (≈ 90 images). The vertical lines identify a new batch of 10 images for illustrative purposes.
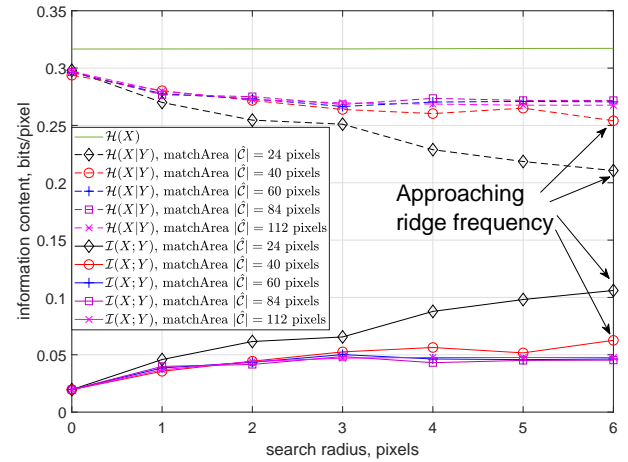

Fig. 9. MI estimates with non-stationary context, taking into account local deformations as a function of the search radius and match area size $|\hat{\mathcal{C}}|$.

local region of the PoI. When the match area size is increased to more than 60 pixels, a convergence can be observed in MI at around search radius of $R = 3$ pixels.

In Fig. 10, the MI per area is given for selected datasets (one per acquisition technology) when local deformations are taken into account. The match area from (10) is set to $\hat{\mathcal{C}} = 112$ pixels. We see that synthetic images exhibit a slight relative improvement (17.5% and 37% in the noisy and noiseless case, respectively). This improvement can be attributed to
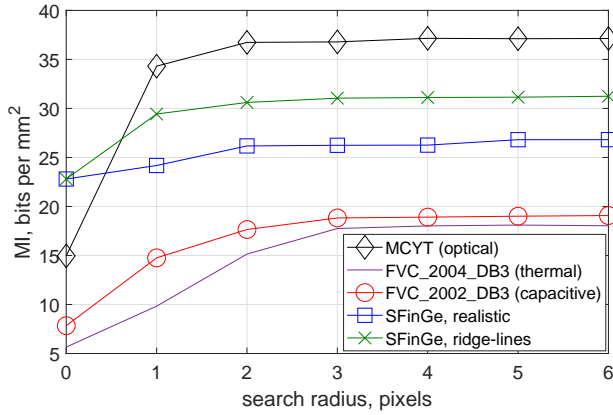
Fig. 10. MI convergence with search radius for the local deformation compensation for selected datasets. Synthetic images converge faster and exhibit smaller relative improvement, which can be attributed to the finite precision of the affine transformation. High-quality optical images take full advantage of the information content of ridge contours and exhibit higher MI.

TABLE III
MI ESTIMATES WITH LOCAL DEFORMATION COMPENSATION.

	MI per pixel	MI per mm^2
FVC-2002-DB1-A	0.044	17.6
FVC-2002-DB2-A	0.034	16.5
FVC-2002-DB3-A	0.047	18.8
FVC-2004-DB1-A	0.038	15.3
FVC-2004-DB2-A	0.051	20.5
FVC-2004-DB3-A	0.045	18.3
MCYT optical	0.091	37.1
MCYT capacitive	0.069	27.9
SFinGe, ridge-lines	0.065	26.0
SFinGe, realistic	0.077	30.8

the imperfections in the affine transformation that is applied in order to estimate the transformed image Y at Step 2 in Algorithm 1. On the other hand, the improvement is more than double for the real datasets. Additionally to the affine transformation imperfections, this improvement is attributed to true local deformations in the images. The MI saturates to the values given in Table III, showing that the MI on true datasets can even exceed that of synthetic images when the local deformations are compensated for due to the high information content of level-3 features (pores and ridge contours), not present in the synthetic images.

An example of the improvement of the conditional entropy is given in the histogram in Fig. 7, where it is shown that more pixels have high (≈ 1) conditional probabilities and their information content thus contributes to the MI.

VI. DISCUSSION AND IMPLICATION OF RESULTS

In this paper, binarized images only were treated due to the complexity, related to higher-order representation. The closed form expression models from Section III-A can also be applied to multi-level images. In such cases, the entropy and MI estimates will clearly be higher, however, the sensitivity to noise is also increased.

The methods for estimating entropy and MI from sequences of data are general to any signal. The presented methodology

of training generative and channel models, then using them for estimation of entropy, MI and thereby identification capacity is thus applicable to other biometric modalities, where closed-form generative and channel models are unavailable.

Single-layer NN was treated in this work in order to showcase the methodology. More sophisticated, deeper network pixel-level models (e.g. as in [27]) can potentially improve the entropy bounds and MI estimates and are interesting area for future research.

As mentioned in Section II, feature extraction is usually performed in order to reduce the dimensionality of the recognition problem. In this paper, raw (binarized) images are treated in order to estimate a capacity with unconstrained recognition algorithms. The identification capacity of systems running practical algorithms with feature extraction can also be estimated with the presented methods. For that purpose, *feature* generative and channel models can be trained. Furthermore, layered models can be devised that first model explicit fingerprint features, such as minutiae, and then tap into the ridge contour information content in a pixel-level fashion. It can be expected that pixel-level MI vastly dominates the information content of such macro features, but this is left for future research to confirm.

It is noted that while all presented results on $\mathcal{H}(X)$ are valid upper bounds, the results on MI can only be considered estimates, since the results on $\mathcal{H}(X|Y)$ are upper bounds themselves. The results on MI are valid as far as either or both upper bounds on $\mathcal{H}(X)$ and $\mathcal{H}(X|Y)$ are tight. We speculate that since the generative and channel models are from the same family and estimated based on the same images X, the tightness and uncertainty of both bounds is similar and the MI estimates are thus reasonable. This is yet to be confirmed by future research.

Biometric entropy and MI are fundamental requirements for estimating bounds on template storage, as well as template security, privacy and invertibility and performance/security trade-offs [28]. The methodology presented in this paper can be applied also to such problems.

Finally, the most conservative MI estimate presented (5.48 bits/mm^2 corresponding to 0.0135 bits/pixel at 512 dpi, FVC-2004-DB3 in Table II) is applied in order to estimate the number of individuals which can be represented with a fingerprint sample. For a sensor of size 80x80 pixels, the result is $2^{80\cdot80\cdot0.0135} \approx 10^{28}$, which is well-above the entire world population that ever lived of $\approx 1.07 \cdot 10^{11}$ [29]. Following the presented analysis, recognition of a population of this size would be possible with a sensor of size $\approx 2.7 \cdot 10^3$ pixels, or ≈ 52x52 (at 512 dpi), as long as the pixel-level generative and channel models hold below the minimum presented size of 80x80 and said population adhere to the generative model. The problem of operating a system at a rate, close to this identification capacity in an efficient manner is still open, and entails as mentioned design of good error-correcting and source codes (or feature vectors). However, from the estimates given in this paper, it can be suggested that when a complexity unconstrained matching algorithm is allowed, small fingerprint sensors will be well-sufficient for identification of large populations. The only requirement is

that the reference image covers sufficient area, such that the overlap between the active probe area and the reference images is sufficiently large.

VII. CONCLUSION

In this paper, the entropy and mutual information (MI) of fingerprint images were estimated. It was demonstrated that texture-based generative models allow for entropy-per-pixel estimation which is independent of the sensor size. The entropy was therefore shown to only depend on the amount of active area, present in the fingerprint sample. By extension, the biometric performance of complexity-unconstrained fingerprint recognition algorithms will only be a function of the overlapping area between the probe and reference samples. The achievable rates of fingerprint identification systems were then estimated via the MI between a probe and reference for several public databases. The most conservative texture-based MI estimates given in this paper suggest capacity, which is sufficient for identification of extremely large populations with sensors as small as a few thousand pixels.

APPENDIX: LIST OF ABBREVIATIONS AND MAIN DEFINITIONS

FVC	*fingerprint verification competition.*
LR	*logistic regression.*
LUT	*look-up table.*
MCYT	*Ministerio de Ciencia y Tecnología.*
MI	*mutual information.*
NN	*neural network.*
OMAP	*orientation map.*
PMF	*probability mass function.*
PoI	*pixel of interest.*
PoO	*probability of occurrence.*
PPM	*pixel predictive model.*
PRC	*probability of random correspondence.*
S	*biometric source.*
X	*biometric source template image after acquisition and processing.*
Y	*biometric source verification image after acquisition and processing.*
$\mathcal{H}(X)$	*information entropy of X.*
$\mathcal{H}(X|Y)$	*conditional entropy of X after observing Y.*
$\mathcal{I}(X;Y)$	*mutual information of X and Y .*

ACKNOWLEDGEMENTS

REFERENCES

[1] R. Cappelli, M. Ferrara, and D. Maltoni, "On the operational quality of fingerprint scanners," *IEEE Transacstions on Information Forensics and Security*, vol. 3, no. 2, pp. 192–202, 2008.

[2] Y. Zhu, S. C. Dass, and A. K. Jain, "Statistical models for assessing the individuality of fingerprints," *IEEE Transacstions on Information Forensics and Security*, vol. 2, no. 3, pp. 391–401, 2007.

[3] M. R. Young, S. J. Elliott, C. J. Tilton, and J. E. Goldman, "Entropy of fingerprints," *IJCSET*, vol. 3, no. 2, pp. 43–47, 2013.

[4] M. B. Westover and J. A. O'Sullivan, "Achievable rates for pattern recognition," *IEEE Transacstions on Information Theory*, vol. 54, no. 1, pp. 299–320, 2008.

[5] S. Pankanti, S. Prabhakar, and A. K. Jain, "On the individuality of fingerprints," *IEEE Transacstions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1010–1025, 2002.

[6] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition, 2nd edition*. Springer, 2009.

[7] A. Roy, N. Memon, and A. Ross, "Masterprint: Exploring the vulnerability of partial fingerprint-based authentication systems," *IEEE Transacstions on Information Forensics and Security*, vol. 12, no. 9, pp. 2013–2025, 2017.

[8] Y. Chen and A. K. Jain, "Beyond minutiae: A fingerprint individuality model with pattern, ridge and pore features," in *Tistarelli M., Nixon M.S. (eds) Advances in Biometrics. ICB 2009. Lecture Notes in Computer Science*, vol. 5558, Springer.

[9] Q. Zhao, D. Zhang, L. Zhang, and N. Luo, "Adaptive fingerprint pore modeling and extraction," *Pattern Recognition*, vol. 43, no. 8, pp. 2833–2844, 2010.

[10] F. Alonso-Fernandez, J. Fierrez, J. Ortega-Garcia, J. Gonzalez-Rodriguez, H. Fronthaler, K. Kollreider, and J. Bigun, "A comparative study of fingerprint image-quality estimation methods," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 4, pp. 734–743, 2007.

[11] R. Youmaran and A. Adler, "Measuring biometric sample quality in terms of biometric information," in *Biometrics Symposium*, Sep. 2006.

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory, 2nd edition*. John Wiley and Sons, Inc, 2006.

[13] F. Willems, T. Kalker, J. Goseling, and J.-P. Linnartz, "On the capacity of a biometrical identification system," in *International Symposium on Information Theory, ISIT*, July 2003.

[14] D. M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Transacstions on Information Theory*, vol. 52, no. 8, pp. 3498–3508, 2006.

[15] C. Su and S. N. Srihari, "Generative models for fingerprint individuality using ridge models," in *International Conference on Pattern Recognition, ICPR*, Dec. 2008.

[16] J. Daugman, "Information theory and the IrisCode," *IEEE Transacstions on Information Forensics and Security*, vol. 11, no. 2, pp. 400–409, 2016.

[17] J. Justensen and S. Forchhammer, *Two-Dimensional Information Theory and Coding*. Cambridge University Press, 2010.

[18] L. Liu, T. Jiang, J. Yang, and C. Zhu, "Fingerprint registration by maximization of mutual information," *IEEE Transactions on Image Processing*, vol. 15, no. 5, pp. 1100–1110, 2006.

[19] "Information technology - coded representation of picture and audio information - progressive bi-level image compression," *ITU Recommendation T.82 (03/93)*, 1993.

[20] P. G. Howard, F. Kossentini, B. Martins, S. Forchhammer, and W. J. Rucklidge, "The emerging JBIG2 standard," *IEEE Transacstions on Circuits and Systems for Video Tech.*, vol. 8, no. 7, pp. 838–848, 1998.

[21] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.

[22] R. Fan, K. Chang, C. Hsieh, X. Wang, and C. Lin, "LIBLINEAR: A library for large linear classification," *Journal of Machine Learning Research*, vol. 9, no. 8, pp. 1871–1874, 2008.

[23] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J. J. Igarza, C. Vivaracho, D. Escudero, and Q. I. Moro, "Mcyt baseline corpus: a bimodal biometric database," *IEE Proceedings on Visual Image Signal Processing*, vol. 150, no. 6, pp. 395–401, 2003.

[24] "Biometric recognition group - ATVS." http://atvs.ii.uam.es/atvs/mcyt100f.html.

[25] "Biometric system laboratory - SFinGe." http://biolab.csr.unibo.it/Research.asp.

[26] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds, "Sheep, goats, lambs and wolves: A statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation," in *International Conference on Spoken Language Processing, ICSLP*, Dec. 1998.

[27] K. Bousmalis, N. Silberman, Dohan, D. Erhan, and D. Krishnan, "Unsupervised pixel level domain adaptation with generative adversarial networks," in *Conference on Computer Vision and Pattern Recognition, CVPR*, July 2017.

[28] K. Nandakumar and A. K. Jain, "Biometric template protection: Bridging the performance gap between theory and practice," *IEEE Signal Processing Magazine*, vol. 32, no. 5, pp. 88–100, 2015.

[29] C. Haub, "How many people have ever lived on earth?," *A Publication of the Population Reference Bureau*, vol. 30, no. 8, pp. 3–4, 2002.