RATE-CONSTRAINED DISTRIBUTED DISTANCE TESTING AND ITS APPLICATIONS

Chuohao Yeo[‡], Parvez Ahammad[§], Hao Zhang[‡], and Kannan Ramchandran[‡]

[‡]Dept. of EECS
University of California, Berkeley
Berkeley, CA 94720, USA

[§]Janelia Farm Research Campus
Howard Hughes Medical Institute
Ashburn, VA 20147, USA

ABSTRACT

We investigate a practical approach to solving one instantiation of a distributed hypothesis testing problem under severe rate constraints that shows up in a wide variety of applications such as camera calibration, biometric authentication and video hashing: given two distributed continuous-valued random sources, determine if they satisfy a certain Euclidean distance criterion. We show a way to convert the problem from continuous-valued to binary-valued using binarized random projections and obtain rate savings by applying a linear syndrome code. In finding visual correspondences, our approach uses just 49% of the rate of scalar quantization to achieve the same level of retrieval performance. To perform video hashing, our approach requires only a hash rate of 0.0142 bpp to identify corresponding groups of pictures correctly.

Index Terms— random projections, distributed hypothesis testing, camera calibration, video hashing

1. INTRODUCTION

The following problem is one that arises in seemingly disparate areas. Suppose there are two distributed sources, one that outputs $\vec{x} \in \mathbb{R}^N$, and another that outputs $\vec{y} \in \mathbb{R}^N$, such that $\|\vec{x}\| = \|\vec{y}\| = 1$[1]. Say Alice observes $\vec{x}$ and Bob observes $\vec{y}$. Under severe rate constraints in a distributed setup, Alice would like to know with high probability if $\|\vec{x} - \vec{y}\|_2 < \tau$; we will refer to this problem as *distributed distance testing under severe rate-constraints*. One solution is for Bob to send some suitably quantized version of $\vec{y}$ to Alice. However, under severe rate constraints in a distributed setup, this might not be suitable. In this paper, we propose a method which uses binarized random projections and linear codes.

One application where such a problem needs to be solved is that of determining visual correspondences in a distributed fashion between cameras in a wireless camera network [1, 2, 3]. This is a critical step for computer vision tasks such as camera calibration, novel view rendering, object recognition and scene understanding. It is usually performed by first locating *features* in input images, computing *descriptors* for each of the features, and then checking the distances between descriptors of features across cameras. In a centralized setting, advances from the computer vision community in locating features and computing descriptors [4, 5] have made establishing visual correspondences reasonably successful. In a distributed camera network setting, the communication costs of exchanging information would need to be accounted for.

Another class of applications where this problem arises is that of image authentication (see for example [6, 7]) and video hashing. Suppose Alice wants to verify that her copy of an image (or video) is similar to what Bob has. Bob can send a perceptual hash of his image to Alice. One approach is to use the actual images such that the transmitted hash should check out if the images satisfy some mean square error (MSE) distortion constraint [7]. Alternatively, we can perform such a distance check in the space of image features [6].

Our main contributions are the following. First, we show that given a euclidean distance threshold, we can perform this test under severe rate constraints by using binarized random projections and linear codes. Second, we show a systematic way of constructing a statistical test and computing the rate of the linear code based on the specified euclidean distance threshold. Third, we show a simple error bound that can be used to determine the number of random projections needed to satisfy an acceptable probability of error.

The rest of the paper is organized as follows. We first review relevant work in Section 2. Sections 3 and 4 discusses the binarized random projections and application of the linear code respectively. We summarize the entire procedure in Section 5, and present experimental results for both visual correspondences and video hashing in Section 6 before concluding.

2. RELATED WORK

Han and Amari presented a survey of work on statistical inference with consideration of communications costs [8]; while they presented theoretical and asymptotic results on achievable error-exponents, no constructive and practical scheme is given. To determine correspondences, Cheng *et al.* introduced a feature digest which applies Principal Components Analysis (PCA) *at each camera* on feature descriptors and then sends only the top principal components [1]. Yeo *et al.* exploited the correlation between descriptors of features in correspondence for rate savings by using distributed source coding (DSC) [2],

[1]In fact, all that is required is that $\vec{x}$ and $\vec{y}$ have the same norm, but for clarity of discussion, we will assume that they have unit norm.

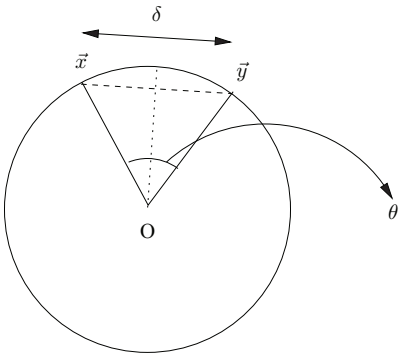

Fig. 1. **Graphical illustration of proof for Fact 1.** A general multi-dimensional case can always be reduced to a 2-D case, in the plane formed by $\vec{x}$, $\vec{y}$, and the origin. The angle subtended by the rays from the origin to $\vec{x}$ and $\vec{y}$ in this plane can be found using simple trigonometry to be $\theta = 2\sin^{-1}(\delta/2)$. If a hyperplane orientation is chosen uniformly at random, then the probability of the hyperplane separating $\vec{x}$ and $\vec{y}$ is just θ/π.

while prescribing bit allocation based on descriptor statistics.

Roy and Sun used binarized random projections to build a descriptor hash [6]; the Hamming distance between hash bits is then used to establish matching features. The statistical link between the original descriptor space and the descriptor hash space was shown empirically, and was further analyzed by Yeo *et al.* [3]. Martinian *et al.* proposed a way of storing biometrics securely using a syndrome code to encode the enrolled biometric bits [9], while Lin *et al.* proposed the use of syndrome codes on quantized projections for image authentication [7]. In both approaches, the syndrome is decoded using the test biometric or test image as side-information; a match is signaled by decoding success. However, the rate of the syndrome code has to be chosen by trial and error to balance security, false positive and false negative performance.

3. BINARIZED RANDOM PROJECTIONS

In previous work, Yeo *et al.* showed the following [3]:

Fact 1. *Given $\vec{x}, \vec{y} \in \mathbb{R}^N$ s.t. $\|\vec{x}\| = \|\vec{y}\| = 1$ and $\|\vec{x}-\vec{y}\|_2 = \delta$, the probability that a randomly (uniformly) generated hyperplane will separate them is $\rho(\delta) = \frac{2}{\pi}\sin^{-1}\frac{\delta}{2}$.*

Proof of Fact 1. We first reduce the problem to a 2-D case as follows. $\vec{x}$, $\vec{y}$ and the origin defines a plane, $\mathcal{S}$. Observe that a hyperplane H passing through the origin separates $\vec{x}$ and $\vec{y}$ *if and only if* the line intersection between H and $\mathcal{S}$ also separates the projections of $\vec{x}$ and $\vec{y}$ on $\mathcal{S}$. The result then follows from trigonometry, as shown in Fig. 1. □

Using Fact 1, we convert the distance testing problem from a deterministic and continuous-valued problem to a probabilistic and binary-valued one. Let $\vec{l_i} \in \mathbb{R}^N$ be some randomly generated vector; this is equivalent to choosing a random hyperplane. Define the following random variables: $X_i = \mathbb{I}\left[\vec{l_i}\cdot\vec{x} > 0\right]$ and $Y_i = \mathbb{I}\left[\vec{l_i}\cdot\vec{y} > 0\right]$. From Fact 1, if $\|\vec{x}-\vec{y}\|_2 = \delta$, then $P(X_i \oplus Y_i = 1) = \rho(\delta)$, since $(X_i \oplus Y_i)$ is 1 *if and only if* the hyperplane separates $\vec{x}$ and $\vec{y}$. Hence, we can model X_i and Y_i as being related by a binary symmetric channel (BSC) with parameter $\rho(\delta)$ when $\|\vec{x}-\vec{y}\|_2 = \delta$.

Denote $\vec{X}, \vec{Y}$ to be binary-valued M-tuples formed by stacking $\{X_i\}_{i=1}^M$ and $\{Y_i\}_{i=1}^M$ respectively, where M is the number of projections taken. The hamming distance between $\vec{X}$ and $\vec{Y}$, $d_H(\vec{X},\vec{Y})$, follows the binomial distribution and can be used as a test statistic in a hypothesis testing framework to decide if $\vec{x}$ and $\vec{y}$ satisfy the distance criterion. Let p denote the probability of a randomly generated hyperplane separating $\vec{x}$ and $\vec{y}$, and let $p_\tau = \rho(\tau)$. The hypotheses are:

$$H_0: \quad p > p_\tau + \mu/2 \quad (\text{i.e. } \|\vec{x}-\vec{y}\| > \tau)$$
$$H_1: \quad p < p_\tau - \mu/2 \quad (\text{i.e. } \|\vec{x}-\vec{y}\| < \tau)$$

where μ specifies an "insensitive" region around p_τ for which we would not measure performance. Since $d_H(\vec{X},\vec{Y})$ has a binomial distribution, it is a monotone likelihood ratio (MLR) statistic [10]. Therefore, we can construct a uniformly most powerful (UMP) test of level α based on thresholding $d_H(\vec{X},\vec{Y})$ with the following properties: the probability of falsely declaring a pair satisfying the distance criterion is always less than α, while the probability of missing a pair satisfying the distance criterion is not more than any other tests of level α [10]. One reasonable choice for the threshold is $\gamma_M = M \cdot p_\tau$. Such an approach has been shown to out-perform scalar quantization in retrieving visual correspondences in the low-bitrate regime [3].

To understand how many projections are needed for a test to satisfy a given error bound, we apply a Chernoff bound on the probability of false detection (declaring H_1 given H_0) and missed detection (declaring H_0 given H_1) of the hypothesis test. For example, given that $p > p_\tau + \mu/2$ (*i.e.* H_0),

$$P(\hat{H}_1|p, H_0) \leq \exp\left(-MD(p_\tau||p)\right) \quad (1)$$
$$\leq \exp\left(-MD(p_\tau||p_\tau+\mu/2)\right) \quad (2)$$

where $D(p||q)$ is the Kullback-Leibler divergence between two Bernoulli sources with parameter p and q, (1) follows from applying Chernoff bound, and (2) follows from considering the worst case in H_0, which is when $p = p_\tau + \mu/2$. In this analysis, we assume the choice of threshold $\gamma_M = Mp_\tau$. A similar analysis can also show that $P(\hat{H}_0|H_1) \leq \exp\left(-MD(p_\tau||p_\tau-\mu/2)\right)$. These bounds can then be used to determine a suitable number of projections to use given a desired error bound.

4. LDPC CODES

In a related work, Körner and Marton [11] showed that if $\vec{X}$ and $\vec{Y}$ are generated by binary symmetric sources related by a BSC with *known* cross-over probability p, then to recover the flip pattern, $\vec{Z} = \vec{X} \oplus \vec{Y}$, with probability of failure less than ϵ, both Alice and Bob need to use at least $H(p)$ bits respectively (asymptotically). The achievable strategy uses a linear code

and is as follows [11]: Let $f(\vec{Z})$ be a linear encoding function (returning K output bits from M input bits) of the binary vector $\vec{Z}$, and $\psi(\cdot)$ be the decoding function of this linear code, such that $P\left(\psi(f(\vec{Z})) \neq \vec{Z}\right) < \epsilon$. Alice and Bob then construct and transmit $f(\vec{X})$ and $f(\vec{Y})$ respectively. The decoder, Alice, can then construct $f(\vec{X}) \oplus f(\vec{Y}) = f(\vec{X} \oplus \vec{Y}) = f(\vec{Z})$, since $f(\cdot)$ is a linear code, and reconstruct $\vec{Z}$ with probability of failure less than ϵ.

While the above scheme recovers the flip pattern $\vec{Z}$, Ahlswede and Csiszár showed that the above rate region in fact holds even if only the hamming distance is desired and $\vec{Y}$ is known at the decoder [12] (and p is known). This also suggests that if we want to recover the hamming distance only when $p < p_\tau$ (but p is otherwise unknown), the best we can hope to do in a one-shot scenario (*i.e.* Bob just sends one message to Alice with no other interaction) is to use a rate of $H(p_\tau)$, and the method described earlier is an achievable strategy. The optimality of this scheme when we just want to know if the hamming distance is smaller than some threshold is an open question.

For a practical implementation used in this work, we use the parity-check matrix of a low-density parity-check (LDPC) code [13] as the linear encoding function [7, 9]; thus, the output $f(\vec{X})$ is just the LDPC syndrome of $\vec{X}$. To decode, we apply belief-propagation (BP) decoding [14] on the XOR sum of the syndromes of $\vec{X}$ and $\vec{Y}$, *i.e.* $f(\vec{X}) \oplus f(\vec{Y})$. We choose a code with blocklength M and rate r such that it has a threshold corresponding to $\frac{\gamma_M}{M}$ [14]. To determine if the distance criterion is satisfied, decoding must converge[2], and the hamming weight of $\vec{Z}$ is less than γ_M.

5. METHOD

The procedure for performing distributed distance testing is as follows. The user parameters are: N, the dimensionality of the real-valued source; M, the number of projections desired; and τ, the euclidean distance threshold (or equivalently $\gamma_M = M\rho(\tau)$). From these parameters, we generate a suitable LDPC code with K syndrome bits, *i.e.* rate $(1 - \frac{K}{M})$, such that it has threshold $\frac{\gamma_M}{M}$, and obtain its parity check matrix $H \in GF(2)^{M \times K}$. We also generate a random projection matrix $L \in \mathbb{R}^{N \times M}$ with the ith column denoted by $\vec{l}_i$. Both H and L are shared by the encoder and decoder.

The encoder takes a vector $\vec{x} \in \mathbb{R}^N$ as input, and returns a binary vector $\vec{m}_x \in GF(2)^K$. It performs the following: (i) Compute the binary random projections, $\vec{X}$, with the ith element being $X_i = \mathbb{I}\left[\vec{l}_i \cdot \vec{x} > 0\right]$; and (ii) Compute the syndrome of $\vec{X}$, $\vec{m}_x = H^T\vec{X}$. The decoder takes two binary vectors, $\vec{m}_x, \vec{m}_y \in GF(2)^K$ ($\vec{m}_y$ is obtained from $\vec{y}$ using the same encoder as described above), and returns H_1 if the distance criterion is satisfied by $\vec{x}$ and $\vec{y}$, and H_0 otherwise. The process is: (i) Compute $\vec{m}_z = \vec{m}_x \oplus \vec{m}_y$; (ii) Perform BP decoding on the syndrome $\vec{m}_z$ to obtain reconstruction $\hat{Z} \in GF(2)^M$; and (iii) If BP decoding converges and $d_H(\hat{Z}) \leq \gamma_M$, return H_1; else, return H_0.

[2]We determine that it converges if the reconstruction satisfies the parity check matrix within 50 iterations.

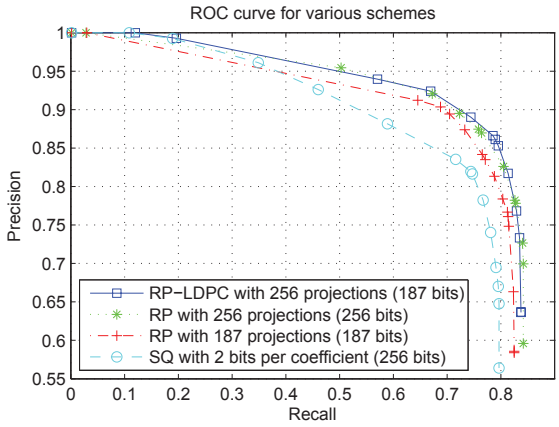

Fig. 2. **ROC curves of determining visual correspondences.**

6. EXPERIMENTAL RESULTS

6.1. Establishing visual correspondences

We first evaluate the performance of three schemes on the problem of visual correspondences retrieval between two cameras: (i) the proposed Random projections with LDPC (RP-LDPC); (ii) Random projections (RP); and (iii) Scalar quantization (SQ). We use the "Graf" dataset made publicly available[3] by Mikolajczyk and Schmid [5], in which images are various views taken of a planar scene. We use two views and extract 1000 features from each image using the Hessian-Affine region detector [5], and compute descriptors of each feature using the 128-dimensional Scale-Invariant Feature Transform (SIFT) descriptor [4]. We note here that SIFT descriptors are normalized in the last step of computation to be robust to illumination changes and thus satisfy the unit-norm assumption. From training data, we determined that a reasonable distance criterion has $\tau = 0.4367$, with $\rho(\tau) = 0.1401$. Accordingly, we use a rate (1-0.73) LDPC code such that the empirically determined threshold of the code is about 0.1401.

Fig. 2 shows the ROC of RP-LDPC with 256 projections, which requires 187 bits per descriptor. Comparing it with the ROC of RP with 256 projections (i.e. 256 bits per descriptor), it is clear that RP-LDPC is able to match the performance of RP using the same number of projections but with less rate. RP-LDPC also outperforms both RP with 187 projections and SQ with 2 bits per coefficient (i.e. 256 bits per descriptor). Compared to SQ with 3 bits per coefficient (i.e. 384 bits per descriptor), RP-LDPC with 187 bits per descriptor has the same level of retrieval performance, but uses only 49% of the former's rate.

[3]`http://www.robots.ox.ac.uk/~vgg/research/affine`

6.2. Video hashing

In a video file synchronization application, video hashing can be used to first determine which group of pictures (GOP) are in common between the source and destination videos [15]. For example, Alice has a video which she gives to Bob who compresses it for storage. Later, Alice updates her copy of the video, and Bob wishes to synchronize his copy. To avoid sending frames that Bob already has, she wishes to know which frames of Bob are within a target distortion of video frames of her copy — these frames need not be re-transmitted.

To demonstrate the effectiveness of using binarized random projections with linear code, we carry out the following experiment. We use the "Foreman" video in QCIF format (176×144 pixels). Our setup is such that Alice has the original video, and Bob has a compressed version of the video, with a PSNR of 42.1 dB. Therefore, the hash should identify that corresponding GOPs from the two videos match. Our target is to identify GOPs with a distortion such that their MSE is less than 10.3 (*i.e.* PSNR is greater than 38 dB).

Consider two blocks of pixels, $\vec{x}$ and $\vec{y}$, with zero means. We assume that they have similar second moments, such that $\|\vec{x}\|^2 = \|\vec{y}\|^2 = N\sigma^2$, where N is the number of pixels in the block. The MSE between the blocks is $\|\vec{x} - \vec{y}\|^2/N = \sigma^2\|\tilde{x} - \tilde{y}\|^2$, where $\tilde{x} = \vec{x}/\sqrt{N\sigma^2}$ and $\tilde{y} = \vec{y}/\sqrt{N\sigma^2}$; thus, $\|\tilde{x} - \tilde{y}\| = \sqrt{MSE/\sigma^2}$. From Fact 1, we then compute $\rho = \frac{2}{\pi}\sin^{-1}\sqrt{MSE/4\sigma^2}$. With our target MSE criterion, and estimating $\sigma^2 = 2940$ from the video, we arrive at a flip probability of $\rho = 0.0188$ and thus a threshold of $\gamma_M = 448$.

The hash is constructed over a GOP of 15 frames as follows [15]. Each frame is divided into non-overlapping blocks of 8×8 pixels. For each block, we subtract the mean pixel value of the entire video, and compute and binarize 4 random projections. Since we want to efficiently check each non-overlapping GOP from Bob with all *overlapping* GOPs from Alice, by using the same set of projection matrices for each frame, we only have to perform the projections on each frame once and use it for all subsequent checks. For each GOP, we first construct a hash of 23760 bits. We then use a rate (1-0.2279) LDPC code to meet this threshold, hence we only need to transmit 5415 syndrome bits per 15 frames, which corresponds to a rate of 0.0142 bits/pixel.

On this particular task, we identified all the matching GOPs correctly without returning any false positives, *i.e.* both recall and precision are 100%. We also note that the same performance is obtained when the LDPC code is not used, *i.e.* there is a rate reduction of 77% with no loss in matching performance when LDPC code is used on top of the binarized random projections.

7. CONCLUDING REMARKS

We have presented a constructive solution for determining in a distributed fashion and under severe rate constraints if two normalized real vectors satisfy a given Euclidean distance criterion. By using binarized random projections, we can convert the problem into a binary hypothesis testing problem, and obtain rate savings by applying a linear code to the computed bits. The rate to use for the code can be easily determined by the desired Euclidean distance threshold. Our experimental results for the two applications of establishing visual correspondences and video hashing show that the proposed method vastly out-performs scalar quantization at low rates.

In future work, we would like to remove the same norm constraints and consider other useful source vector distributions and distance measures. We have not explored any security properties of our scheme, but we think that the proposed scheme offers some inherent security, due to the data obfuscation performed by both the binarized random projections and the syndrome coding [9].

8. REFERENCES

[1] Zhaolin Cheng, Dhanya Devarajan, and Richard J. Radke, "Determining vision graphs for distributed camera networks using feature digests," *EURASIP Journal on Advances in Signal Processing*, 2007.

[2] Chuohao Yeo, Parvez Ahammad, and Kannan Ramchandran, "A rate-efficient approach for establishing visual correspondences via distributed source coding," in *Proc. SPIE Visual Communications and Image Processing*, Jan 2008.

[3] Chuohao Yeo, Parvez Ahammad, and Kannan Ramchandran, "Rate-efficient visual correspondences using random projections," in *Proc. IEEE International Conference on Image Processing*, Oct 2008.

[4] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[5] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

[6] Sujoy Roy and Qibin Sun, "Robust hash for detecting and localizing image tampering," in *Proc. IEEE International Conference on Image Processing*, Sep 2007.

[7] Y.C. Lin, D. Varodayan, and B. Girod, "Image Authentication based on Distributed Source Coding," in *Proc. IEEE International Conference on Image Processing*, Sep 2007.

[8] T.S. Han and S. Amari, "Statistical inference under multiterminal data compression," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2300–2324, 1998.

[9] E. Martinian, S. Yekhanin, and J.S. Yedidia, "Secure Biometrics Via Syndromes," in *Allerton Conference on Communications, Control and Computing*, Sep 2005.

[10] P.J. Bickel and K.A. Doksum, *Mathematical statistics: basic ideas and selected topics. Vol. 1*, Prentice Hall, 2 edition, 2000.

[11] J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources (Corresp.)," *IEEE Transactions on Information Theory*, vol. 25, no. 2, pp. 219–221, 1979.

[12] R. Ahlswede and I. Csiszár, "To get a bit of information may be as hard as to get full information," *IEEE Transactions on Information Theory*, vol. 27, no. 4, pp. 398–408, 1981.

[13] R G Gallager, "Low-Density Parity-Check Codes," *MIT Press*, 1963.

[14] T.J. Richardson and R.L. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 599–618, 2001.

[15] Hao Zhang, Chuohao Yeo, and Kannan Ramchandran, "Vsync — a novel video file synchronization protocol," in *Proc. ACM International Conference on Multimedia (ACM MM)*, Oct 2008.