Redundant Metering for Integrity with Information-Theoretic Confidentiality

David P. Varodayan
Department of Electrical Engineering
Stanford University
varodayan@stanford.edu

Grace Xingxin Gao
Department of Aeronautics and Astronautics
Stanford University
gracegao@stanford.edu

***Abstract*—Redundant metering is frequently used to verify the integrity of billing data reported by advanced metering infrastructure, but the redundant measurement introduces a potential confidentiality leak. We propose a way to encode the redundant measurement at a bit rate below its entropy, so that it cannot be decoded from the encoded bits alone. In this way, we guarantee information-theoretic confidentiality, regardless of the computational power of an eavesdropper. We provide practical Slepian-Wolf codes to realize security of up to 5 bit/sample for 8-bit samples based on actual power metering experiments.**

I. Introduction

Advanced metering infrastructure promises a future of highly efficient energy resource usage by enabling services like demand response and time-of-use pricing, but the advances must not compromise the main function of metering: billing. This paper considers the integrity of billing from the customer's point of view. We combine redundant metering with a method to guarantee data confidentiality in an information-theoretic (rather than computational) sense.

The integrity of smart meter data is a paramount concern for customers (as well as utilities). A lack of confidence may lead to a public and political backlash, like the one against Pacific Gas and Electricity (PG&E) in Bakersfield, California in 2009. Thousands of customers complained that their smart meters were overcharging them. PG&E attributed the higher bills to a rate increase and spikes in usage due to summer weather [1]. However, in at least some cases, it appears that billing discrepancies were due to improper installation or malfunctioning equipment [2]. The consequences for PG&E are ongoing lawsuits, political pressure for a moratorium on deployment and increased scrutiny of their smart meters [3], [4]. Some PG&E customers now verify the integrity of their billing independently using consumer-device wireless power meters [5]. The overall setup is shown in Fig. 1. The advanced metering infrastructure provides a direct measurement of the customer's power use to the utility, which relays a reported measurement back to the customer. The customer independently monitors power usage, recording a redundant measurement.

But redundant wireless power metering introduces a potential confidentiality leak. An eavesdropper could easily discover the instantaneous power usage of the customer's home and, hence, whether it is occupied or not. We develop a method for

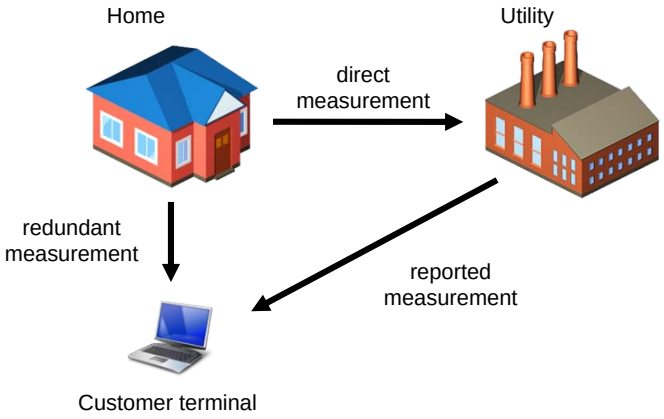

Fig. 1. Advanced metering infrastructure provides the utility with a direct measurement of the customer's power usage. The utility relays this information as a reported measurement to the customer's terminal. The customer monitors power usage independently, recording a redundant measurement. This redundant measurement verifies the integrity of the reported measurement, but introduces a potential confidentiality leak for the customer.

information-theoretic confidentiality of the redundant meter data. The key idea is to compress the redundant measurement to a rate *below* its entropy, so that it cannot be recovered from just the encoded bits. But the redundant measurement can be recovered in conjunction with the reported measurement, as long as the compression rate is greater than the conditional entropy of the redundant measurement given the reported measurement. Unlike encryption, this method guarantees confidentiality regardless of the computational capability of the eavesdropper [6].

Section II breaks down information-theoretic confidentiality as a combination of the Shannon source coding theorem [7] and the Slepian-Wolf theorem [8]. In Section III, we describe practical Slepian-Wolf coding techniques for binary and multilevel signals. We then present a case study of information-theoretic confidentiality for redundant meter data from the Stanford PowerNet project [9] in Section IV. Section V addresses some limitations and future work.

II. Information-Theoretic Confidentiality

Let X and Y represent the redundant and reported measurements, respectively. Recall that we are interested in guaranteeing the confidentiality of X regardless of the computational

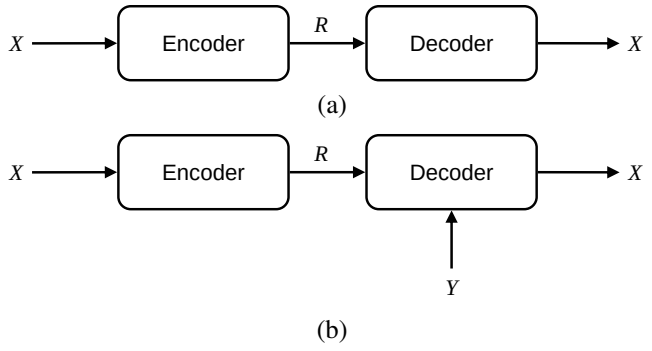

Fig. 2. Compression of X (a) without statistically dependent side information and (b) with statistically dependent side information Y at the decoder only.

capability of an eavesdropper that only knows the encoded version of X. To understand how we can achieve this, we combine two theorems on source coding.

Fig. 2(a) depicts the lossless source coding of X. Shannon's theorem states that lossless recovery is possible if the coding rate $R > H(X)$, the entropy of X; conversely, if $R < H(X)$, then the probability of error in the recovered signal will be bounded away from zero [7].

Fig. 2(b), in contrast, represents lossless source coding of X, where a statistically dependent signal Y is available at the decoder only. Slepian and Wolf show the surprising result that we can encode X to as low a rate as if Y were available at the encoder as well; that is, lossless recovery is possible if $R > H(X|Y)$, the conditional entropy of X given Y [8]. Note that $H(X|Y) < H(X)$ if Y is indeed statistically dependent on X.

Putting these two results together in the following way gives rise to a notion of information-theoretic confidentiality. We encode the redundant measurement X to a rate R such that $H(X) < R < H(X|Y)$. Thus, an eavesdropper that only receives the encoded bits at rate R cannot recover X. But the customer, with the reported measurement Y available, can decode X.

Slepian-Wolf coding in this fashion has found application to information-theoretic security of both biometrics [10] and multimedia [11].

III. Practical Slepian-Wolf Coding

We now describe the practical Slepian-Wolf coding of binary signals with low-density parity-check (LDPC) codes [12] and then extend the technique to the coding of multilevel signals.

Fig. 3 shows an example bipartite encoding/decoding graph of an LDPC code for Slepian-Wolf coding of binary signals. The encoding works as follows. The source bits of X are placed in the bit nodes of the graph. Syndrome bits S are computed at the syndrome nodes as the modulo 2 sum of the connected source bits. The bit rate R is the ratio of number of syndrome nodes to number of bit nodes. The decoder seeks to recover X using the syndrome bits S. In the absence of Y, this is not possible. But, in the presence of statistically dependent

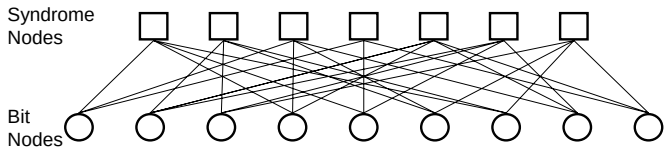

Fig. 3. Bipartite encoding/decoding graph of an LDPC code for Slepian-Wolf coding of binary signals.

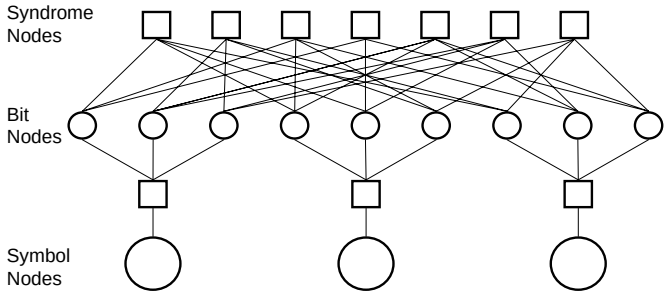

Fig. 4. Augmented decoding graph for Slepian-Wolf coding of multilevel signals.

Y, the decoder seeds the bit nodes with probabilities of the bits of X given the corresponding bits of Y. Then an iterative sum-product algorithm is run until convergence, yielding estimates of the bits of X [13].

We extend this technique to Slepian-Wolf coding of multilevel signals. The multilevel encoder maps the symbols into binary through a Gray code. It then encodes the bits as in the binary case using an encoding graph like the one in Fig. 3. The multilevel decoder, on the other hand, cannot use the same graph for decoding because the multilevel symbols of Y inform the decoder about bits of X only indirectly through the symbols of X. Therefore, the multilevel decoder operates on an augmented decoding graph like the one in Fig. 4. In this example, each symbol is 8-valued and so each symbol node is connected to 3 bit nodes. The decoder seeds the symbol nodes of this graphs with probability mass functions of the symbols of X given the corresponding symbols of Y. Then the decoder runs the sum-product algorithm, presented in detail in [14], until estimates of the symbol values converge.

IV. Case Study: Stanford PowerNet

We validate our information-theoretic confidentiality scheme on data from the Stanford PowerNet project [9].

The instantaneous power draw of a computer (Dell Precision T3400, quad core 2.6 GHz) is measured by two meters. A commercial wired meter provides timestamped power readings in watts at a fixed rate of 1.00 Hz. For our experiment, the wired meter timestamps and power readings constitute the reported measurement. Meanwhile, a wireless meter built at Stanford [15] also provides timestamped power readings, but at a fluctuating rate, averaging 1.06 Hz. The wireless meter timestamps and power readings form the redundant measurement in our experiment. But Slepian-Wolf coding is applied only to the power meter readings. The timestamps,

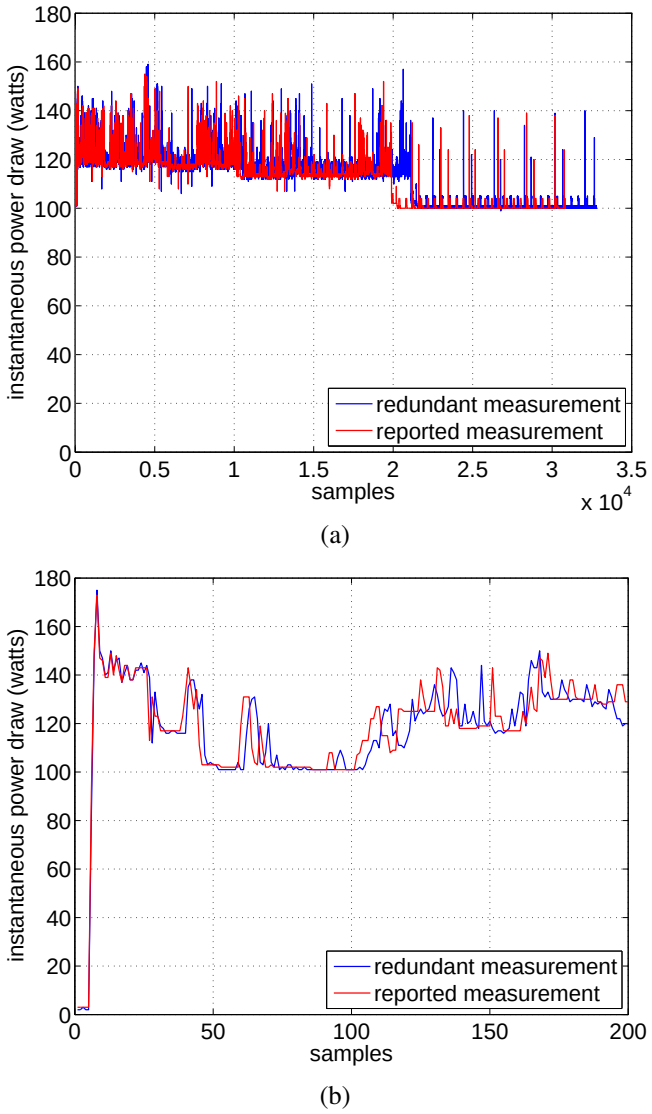

Fig. 5. Redundant and reported measurements in watts for (a) the entirety of the data and (b) the first 200 samples.

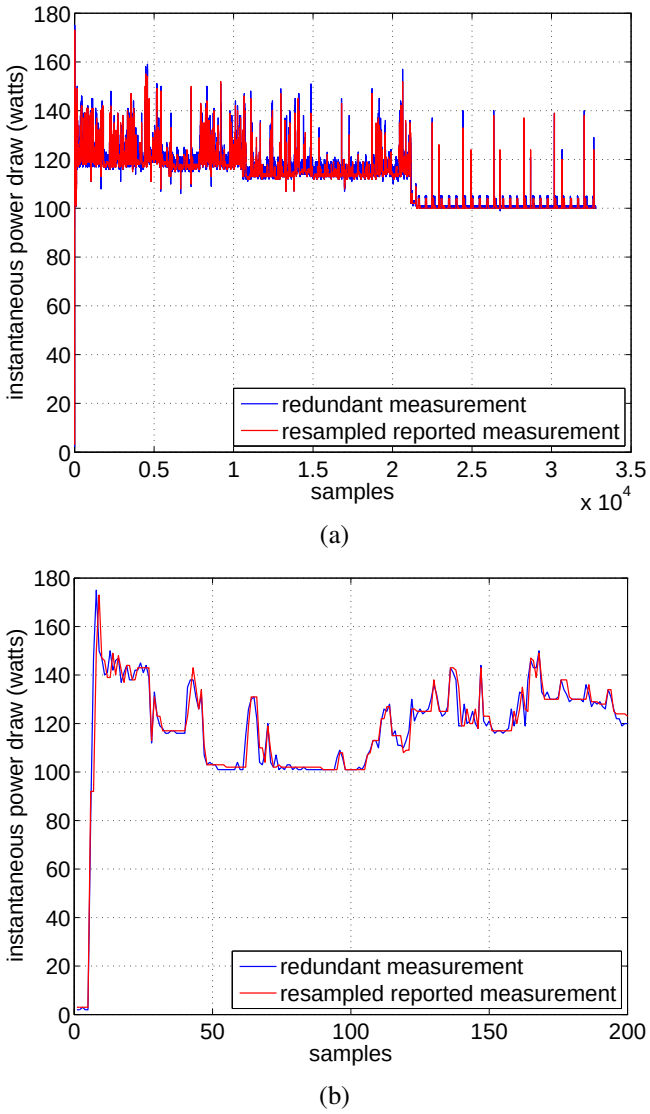

Fig. 6. Redundant and resampled reported measurements in watts for (a) the entirety of the data and (b) the first 200 samples.

which represent less of a confidentiality risk, are sent to the customer's terminal uncoded.

We use data consisting of 30872 reported samples and 32768 redundant samples, collected during almost 9 hours on March 29, 2010. Fig. 5(a) and (b) plot the instantaneous power draw readings (quantized to 8 bits) versus sample number for the entirety of the data and the first 200 samples of each trace, respectively. These figures indicate that, sample by sample, the reported measurement is not good side information for decoding the redundant measurement, because they are not synchronized.

We therefore resample (by nearest neighbor interpolation) the reported measurement so that it aligns with the redundant measurement timestamps provided uncoded by the wireless meter. The resampled reported readings and the redundant readings are shown in Fig. 6(a) and (b) for the entirety of the data and for the first 200 samples, respectively. The resampled reported measurement is now good side information for the decoding of the power readings of the redundant measurement. Fig. 7 shows the probability mass function of the measurement difference over 32768 samples, along with a Laplacian distribution with scale parameter $\beta = 0.8$. Note that the probability mass function is not perfectly centered; that is, the meters are not perfectly calibrated.

We now report the results of the information-theoretic confidentiality method. The 32768 samples of the redundant measurement are divided into 64 blocks of 512 8-bit samples. Each block therefore consists of 4096 bits. We encode each block X of the redundant measurement with a rate-adaptive LDPC codes [16] of length 4096 bits. At the decoder, the corresponding block of the resampled reported measurement is denoted side information Y. In the augmented decoding

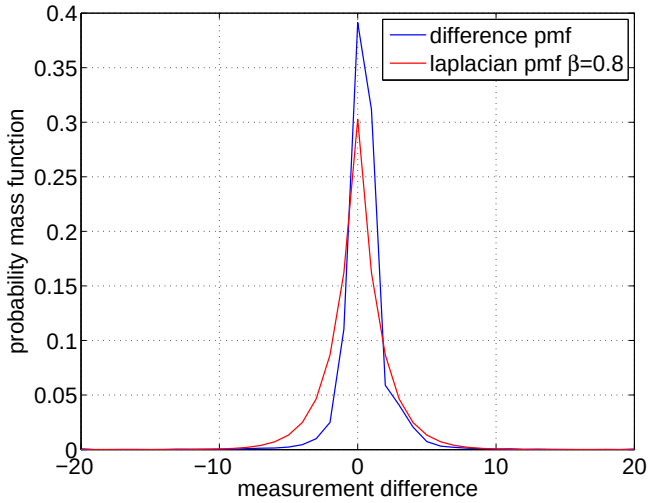

Fig. 7. Probability mass function of the measurement difference over 32768 samples and Laplacian distribution with scale parameter $\beta = 0.8$.

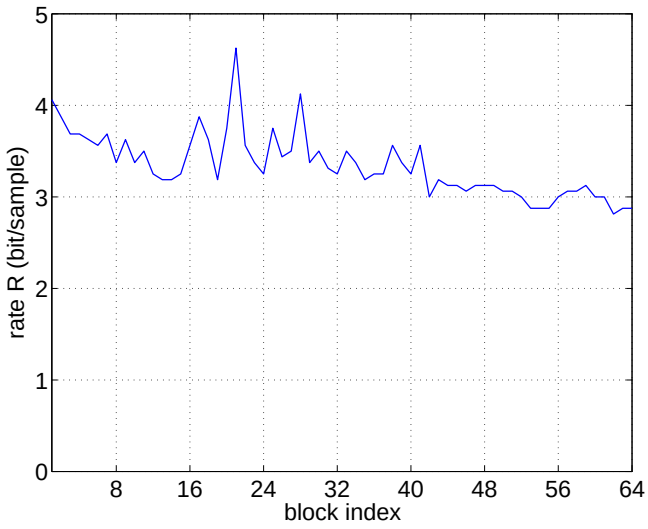

Fig. 8. Minimum Slepian-Wolf coding rate R (in bit/sample) required to recover X in the presence of Y.

graph, each symbol node has degree 8 and is initialized with a Laplacian probability mass function centered at the corresponding sample of Y with scale parameter $\beta = 0.8$. Fig. 8 shows the minimum coding rates R (in bit/sample) required to recover the blocks X in the presence of the blocks Y. At around 3 to 4.5 bit/sample, they are significantly lower than the raw bit rate of 8 bit/sample. The memoryless entropy $H(X)$ for the redundant measurement is close to 8 bit/sample too, because the readings can be arbitrary. Therefore, our method provides about 3.5 to 5 bit/sample of security with respect to the memoryless entropy $H(X)$.

Note that these results assume that the resampled reported measurement and the redundant measurement are statistically in agreement. If they differ significantly, then it is not be possible to decode the redundant measurement at rates R around 3 to 4.5 bit/sample. Such a decoding failure signifies a that the integrity of the reported measurement is unverified.

V. Limitations and Future Work

The treatment of information-theoretic confidentiality so far has ignored the effect of memory in the redundant measurement. Since typical redundant measurements have nontrivial memory, the coding rate R should be chosen in between the entropy rate $\mathcal{H}(X)$ and the conditional entropy rate $\mathcal{H}(X|Y)$ to ensure confidentiality; that is, $\mathcal{H}(X|Y) < R < \mathcal{H}(X)$. But $\mathcal{H}(X)$ and $\mathcal{H}(X|Y)$ are less than $H(X)$ and $H(X|Y)$, respectively, so the R should be lower than under the memoryless assumption. Practical Slepian-Wolf coding can achieve the necessary R by augmenting the decoding graph with a Markov memory model (for example) as in [17].

The Slepian-Wolf encoder and decoder presented in this paper work quite well when the redundant and reported measurements are reasonably (but not necessarily perfectly) calibrated. But they would not work as efficiently if the two meters are significantly uncalibrated. If the calibration parameters are known at the decoder, then it can compensate the resampled reported measurement to approximate the redundant one. If the calibration parameters are not known in advance, then the decoder can use an expectation maximization algorithm [18] to simultaneously estimate the parameters and decode X (see [19]).

VI. Conclusions

In recent deployments of advanced metering infrastructure, customers have frequently questioned the integrity of billing data. Some customers have started independently monitoring their power usage using wireless power meters. This redundant metering verifies the integrity of billing, but also creates a confidentiality risk for the customer. An eavesdropper can easily learn whether the home is occupied or not. In this paper, we propose a method for redundant metering for integrity that also achieves information-theoretic confidentiality. The idea is to compress the redundant measurement below its entropy. Thus, the redundant measurement cannot be recovered from just its encoded bits, and so is information theoretically secure regardless of the computational power of an eavesdropper. The encoded bits, however, can be decoded in the presence of a statistically dependent reported measurement. We provide practical codes for such a scheme and demonstrate its viability in a case study. We suggest improvements of the idea to handle memory in the measurement traces and uncalibrated meters.

Acknowledgment

We thank the members of the Stanford PowerNet project [9], especially Maria Kazandjieva, for sharing their data.

References

[1] Greentech Media, "PG&E Sued Over Smart Meters, Slows Down Bakersfield Deployment," 2009. http://www.greentechmedia.com/articles/read/pge-sued-over-smart-meters-slows-down-bakersfield-deployment.

[2] San Jose Mercury News, "PG&E details technical problems with SmartMeters," 2010. http://www.mercurynews.com/ci_14963541.
[3] Greentech Media, "PG&E to Bakersfield: Weather, Not Smart Meters, Cause of Higher Power Bills," 2009. http://www.greentechmedia.com/green-light/post/pge-to-bakersfield-weather-not-smart-meters-cause-of-higher-power-bills.
[4] KGO-TV, "Regulators unsupportive of SmartMeters moratorium," 2010. http://abclocal.go.com/kgo/story?section=news/7_on_your_side&id=7410033.
[5] KGO-TV, "Experiment raises questions about SmartMeters," 2010. http://abclocal.go.com/kgo/story?section=news/7_on_your_side&id=7424533.
[6] B. Schneier, *Applied Cryptography: protocols, algorithms and source code in C*, John Wiley & Sons, Inc, 1996.
[7] C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. J*, vol. 27, pp. 379–423, 623–656, 1948.
[8] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.
[9] M. Kazandjieva, B. Heller, D. Gal, P. Levis, C. Kozyrakis, and N. McKeown, "PowerNet," 2010. http://powernet.stanford.edu.
[10] E. Martinian, S. Yekhanin, and J. S. Yedidia, "Secure biometrics via syndromes," in *Proc. Allerton Conf. Commun., Contr. and Comput.*, Monticello, Illinois, 2005.
[11] Y.-C. Lin, D. P. Varodayan, and B. Girod, "Image authentication based on distributed source coding," in *Proc. IEEE Internat. Conf. Image Process.*, San Antonio, Texas, Sept. 2007.
[12] A. Liveris, Z. Xiong, and C. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Commun. Lett.*, vol. 6, no. 10, pp. 440–442, Oct. 2002.
[13] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
[14] D. P. Varodayan, D. M. Chen, M. Flierl, and B. Girod, "Wyner-Ziv coding of video with unsupervised motion vector learning," *EURASIP Signal Process.: Image Commun. J*, vol. 23, no. 5, pp. 369–378, Jun. 2008.
[15] M. Kazandjieva, B. Heller, P. Levis, and C. Kozyrakis, "Energy dumpster diving," in *Workshop Power Aware Computing (HotPower)*, Big Sky, Montana, 2009.
[16] D. P. Varodayan, A. M. Aaron, and B. Girod, "Rate-adaptive codes for distributed source coding," *EURASIP Signal Process. J.*, vol. 86, no. 11, pp. 3123–3130, Nov. 2006.
[17] D. P. Varodayan, A. M. Aaron, and B. Girod, "Exploiting spatial correlation in pixel-domain distributed image compression," in *Proc. Picture Coding Symp.*, Beijing, China, 2006.
[18] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc., Series B*, vol. 39, no. 1, pp. 1–38, 1977.
[19] D. P. Varodayan, A. Mavlankar, M. Flierl, and B. Girod, "Distributed coding of random dot stereograms with unsupervised learning of disparity," in *Proc. IEEE Internat. Workshop Multimedia Signal Process.*, Victoria, British Columbia, Canada, 2006.