2004 International Conference on Image Processing (ICIP)

UNIVERSAL MINIMAX BINARY IMAGE DENOISING UNDER CHANNEL UNCERTAINTY

George Gemelos *Styrmir Sigurjónsson* *Tsachy Weissman**

Stanford University
Department of Electrical Engineering
Stanford 94305-9510
{ggemelos, styrmir, tsachy}@stanford.edu

ABSTRACT

We consider the problem of denoising a binary image corrupted by a noisy medium which flips each component in the original image, independently, to its complementary value with some fixed but unknown probability $\delta < 1/2$. We propose a denoiser which assumes no knowledge of statistical properties of the image, yet asymptotically attains the performance of the scheme that knows the noisy image statistics and operates optimally in a minimax sense. The proposed scheme is implementable, with complexity linear in the image size. Preliminary experimental results are presented which indicate that the scheme has the potential to do well on real data.

1. INTRODUCTION

In [1], a practical scheme for recovering a binary image from its noise-corrupted version was presented. The scheme was shown to be universal in the sense of asymptotically, for large images, attaining optimum performance no matter what the statistics of the image turn out to be. This scheme was obtained by generalizing the discrete universal denoiser (DUDE) of [2] to multi-dimensionally indexed data. The experimental results presented in [1] showed that the algorithm outperforms other popular schemes for binary image denoising on various types of binary images.

The assumption of a known channel inherent in the structure of the DUDE of [2] was inherited by the binary image denoiser of [1], which assumes knowledge of the channel crossover probability. This assumption is indeed a realistic one in many practical scenarios where the noisy medium through which the image is observed is well-characterized statistically. Furthermore, even in many applications where this is not the case, the simplicity of the scheme in [1] allows for a practically designed "knob" which, in real time, enables a human observer to subjectively select the reconstruction which looks best among all reconstructions corresponding to employing the scheme of [1] for the whole range of possible noise characteristics. Thus, in practice, the scheme of [1] is effective for denoising under channel uncertainty in applications where it is reasonable to expect availability of "feedback" on the quality of reconstruction.

On the other hand, such feedback on the quality of reconstruction is not realistic in other scenarios, such as those involving processing of large databases of noisy images, or those involving texture images, where human feedback may be too subjective. In such cases, an automated algorithm for image denoising is sought, which will accommodate uncertainty in the statistical characteristics of the noisy medium. With this motivation, the case of uncertainty in the channel characteristics, in addition to the uncertainty in the distribution of the noise-free signal, was recently studied in [3]. It was shown that in this setting the task of attaining the performance of the optimum non-universal Bayesian scheme, which was shown to be attainable in the setting of [2], is impossible, even for a "genie-aided" scheme with complete knowledge of the noisy signal statistics. Under these circumstances, the criterion suggested for judging the performance of a denoising scheme was its worst case performance under all noise-free source distributions and channels consistent with the noisy source distribution. This is a lower bound on attainable performance for the setting of channel uncertainty.

In this work, we present the practical binary image denoiser obtained by extending the scheme of [3] to two-dimensionally indexed data. We start by introducing notation, definitions, and a concrete formulation of the problem in Section 2. Section 3 describes the suggested binary image denoiser, which we show in Section 4 to be asymptotically optimal in the minimax sense that we argue most appropriate for the present setting of noise distribution uncertainty. We discuss the implementation of the scheme and report encouraging preliminary experimental results in Section 5. We conclude in Section 6 by summarizing and mentioning directions in which our scheme and results can be extended.

2. NOTATION AND DEFINITIONS

Throughout we assume the components of the clean, as well as those of the noise-corrupted image, take their values in $\{0, 1\}$. We assume that the noiseless image is corrupted by a binary symmetric channel (BSC) with an unknown crossover probability $\delta < 1/2$. That is to say that each component in the noise-free image is flipped to the complementary value with probability δ, independently of other

*This work was supported by MURI Grant DAAD-19-99-1-0215, NSF Grant CCR-0311633 and in part by NSF Grant CCR-0312839.

components.

A *binary image*, or simply *image*, is a two-dimensional array of $\{0,1\}$-valued components. We let $\mathbf{x} = \{x_i\}_{i\in\mathbb{N}^2}$ denote an infinite noise-free image and $\mathbf{z} = \{z_i\}_{i\in\mathbb{N}^2}$ its noise-corrupted version, where $\mathbb{N}$ is the set of positive integers. For any $\mathcal{S} \subseteq \mathbb{N}^2$ we also denote $x(\mathcal{S}) = \{x_i\}_{i\in\mathcal{S}}$ and $z(\mathcal{S}) = \{z_i\}_{i\in\mathcal{S}}$. Thus $x(\mathcal{S})$ is an $|\mathcal{S}|$-dimensional vector, with $\{0,1\}$-valued components that are indexed by the elements of $\mathcal{S}$. We let $\{0,1\}^{\mathcal{S}}$ denote the set of all such $|\mathcal{S}|$-dimensional vectors. For $m, n \in \mathbb{N}$ let $V_{m\times n}$ denote the $m \times n$ rectangle $V_{m\times n} = \{i = (i_x, i_y) \in \mathbb{N}^2 : i_x \leq m, i_y \leq n\}$. To simplify notation we shall write $x_{m\times n}$ for $x(V_{m\times n})$, and $z_{m\times n}$ for $z(V_{m\times n})$. We will also write $\{0,1\}^{m\times n}$ for $\{0,1\}^{V_{m\times n}}$. For $\mathcal{S} \subseteq \mathbb{Z}^2$ and $i \in \mathbb{N}^2$ we let $\mathcal{S} + i = \{j + i : j \in \mathcal{S}\}$ denote the i'shift of $\mathcal{S}$.

A *randomized* $m \times n$ *binary image denoiser* is a mapping $\hat{X}^{m\times n} : \{0,1\}^{m\times n} \to [0,1]^{m\times n}$. The interpretation is that upon seeing the noisy signal $z_{m\times n} \in \{0,1\}^{m\times n}$, the reconstruction at the i-th location is 1 with probability $\hat{X}^{m\times n}(z_{m\times n})[i]$,where $\hat{X}^{m\times n}(z_{m\times n})[i]$ denotes the component of $\hat{X}^{m\times n}(z_{m\times n})$ at the i-th location. Thus, for $x_{m\times n}, z_{m\times n} \in \{0,1\}^{m\times n}$ we denote

$$L_{\hat{X}^{m\times n}}(x_{m\times n}, z_{m\times n}) = \frac{1}{mn}\sum_{i\in V_{m\times n}} \left| x_i - \hat{X}^{m\times n}(z_{m\times n})[i] \right|,$$

which is the normalized expected number of errors that will be made by the image denoiser $\hat{X}^{m\times n}$ when the observed noisy image is $z_{m\times n}$ and the underlying one is $x_{m\times n}$, the expectation taken with respect to the randomization[1].

For a family of binary image denoisers $\{\hat{X}^{m\times n}\} = \{\hat{X}^{m\times n}\}_{m,n}$ and a binary random field $\mathbf{X} = \{X_i\}_{i\in\mathbb{N}^2}$ distributed according to $P_{\mathbf{X}}$ let

$$L_{\{\hat{X}^{m\times n}\}}(P_{\mathbf{X}}, \delta) = \limsup_{m,n\to\infty} E_{P_{\mathbf{X}},\delta}\left[L_{\hat{X}^{m\times n}}(X_{m\times n}, Z_{m\times n})|\mathbf{Z}\right],$$

where $E_{P_{\mathbf{X}},\delta}$ denotes expectation when $\mathbf{X} \sim P_{\mathbf{X}}$ and $\mathbf{Z}$ is the BSC(δ)-corrupted version of $\mathbf{X}$.

Let $\mathbf{Z} = \{Z_i\}_{i\in\mathbb{N}^2}$ be a spatially stationary binary random field and let $P_{\mathbf{Z}}$ denote its distribution. Denote further

$$\Delta(P_{\mathbf{Z}}) = \max\{0 \leq \delta \leq 1/2 : \exists P_{\mathbf{X}} \text{ s.t. } P_{\mathbf{X}} * \delta = P_{\mathbf{Z}}\}, \quad (1)$$

where $P_{\mathbf{X}} * \delta$ denotes the distribution of a random field distributed according to $P_{\mathbf{X}}$ and corrupted by a BSC(δ). In words, $\Delta(P_{\mathbf{Z}})$ is the largest channel crossover probability consistent with $P_{\mathbf{Z}}$ in the sense of the existence of a field giving rise to this distribution when corrupted by a BSC($\Delta(P_{\mathbf{Z}})$). Define now $\mathbb{MM}(P_{\mathbf{Z}}) =$

$$\min_{\{\hat{X}^{m\times n}\}} \sup_{\{(P_{\mathbf{X}},\delta): P_{\mathbf{X}}*\delta = P_{\mathbf{Z}}, \delta\in[0,\Delta(P_{\mathbf{Z}})]\cap\mathbb{Q}\}} L_{\{\hat{X}^{m\times n}\}}(P_{\mathbf{X}}, \delta), \quad (2)$$

the minimum being over all possible families of sliding window binary image denoisers, i.e. denoisers that employ space invariant filters that depend on a finite number of observations, and $\mathbb{Q}$ denotes the set of rationals. By its definition, $\mathbb{MM}(P_{\mathbf{Z}})$ is an absolute lower bound on the asymptotic performance of any family of sliding window denoisers, even a "genie-aided" one that has access to the noisy image distribution $P_{\mathbf{Z}}$, in the sense that for every such family there exists a clean image source $P_{\mathbf{X}}$ that gives rise to $P_{\mathbf{Z}}$ when corrupted by some BSC(δ), $0 \leq \delta \leq 1/2$, and such that

$$L_{\{\hat{X}^{m\times n}\}}(P_{\mathbf{X}}, \delta) \geq \mathbb{MM}(P_{\mathbf{Z}}) \quad a.s. \quad (3)$$

under that source. We shall thus say that a family of image denoisers is *minimax optimal for the noisy image distribution* $P_{\mathbf{Z}}$ if

$$\mathbb{MM}(P_{\mathbf{Z}}) \geq \sup_{\{(P_{\mathbf{X}},\delta): P_{\mathbf{X}}*\delta = P_{\mathbf{Z}}, \delta\in[0,\Delta(P_{\mathbf{Z}})]\cap\mathbb{Q}\}} L_{\{\hat{X}^{m\times n}\}}(P_{\mathbf{X}}, \delta) \quad P_{\mathbf{Z}} - a.s.$$

A minimax optimal sequence of schemes, even when the distribution $P_{\mathbf{Z}}$ is completely known, is, in general difficult to obtain explicitly. A fortiori, when this distribution is not a priori given, the task may seem even more difficult. Nevertheless, in the next section we describe a practical scheme that assumes no a priori knowledge of the said distribution, and yet is guaranteed to be minimax optimal whatever this distribution may turn out to be.

3. DESCRIPTION OF THE DENOISER

For $\alpha \in [0,1]$, $\delta < 1/2$, $d_0 \in [0,1]$ and $d_1 \in [0,1]$ let

$$F(\alpha, \delta, d_0, d_1) = \frac{1}{1-2\delta}[(1-\delta-\alpha)(1-\delta)d_0 + (1-\delta-\alpha)\delta d_1 + (\alpha-\delta)\delta(1-d_0) + (\alpha-\delta)(1-\delta)(1-d_1)].$$

In the single observation problem, for $\alpha \in [\delta, 1-\delta]$, $F(\alpha, \delta, d_0, d_1)$ is readily verified to be the expected loss of a scheme which says 1 with probability d_0 upon observing a zero and says 1 with probability d_1 upon observing a one; when observing a Bernoulli($(\alpha-\delta)/(1-2\delta)$) corrupted by a BSC($\delta$), so that the channel output is a Bernoulli(α).

Let now $\mathcal{S}$ be a finite subset of $\mathbb{Z}^2$ containing the point $\mathbf{0} = (0,0)$. For a $\{0,1\}^{\mathcal{S}}$-valued random field $U(\mathcal{S})$ let $P_{U(\mathcal{S})}$ denote its distribution. For $f : \{0,1\}^{\mathcal{S}} \to [0,1]$ we define now the functional $G_{\mathcal{S}}$ by

$$G_{\mathcal{S}}\left(P_{U(\mathcal{S})}, \delta, f\right) = \quad (4)$$

$$\sum_{c_{\mathcal{S}}} F\left(P_{U(\mathcal{S})}(U_{\mathbf{0}} = 1|c_{\mathcal{S}}), \delta, f(c_{\mathcal{S}}, 0), f(c_{\mathcal{S}}, 1)\right) P_{U(\mathcal{S})}(c_{\mathcal{S}}),$$

where: 1) The summation is over $c_{\mathcal{S}} \in \{0,1\}^{\mathcal{S}\setminus\mathbf{0}}$, i.e., the set of all $2^{|\mathcal{S}|-1}$ possible "contexts" of $U_{\mathbf{0}}$ that $U(\mathcal{S}\setminus\mathbf{0})$

[1] Though in the setting of [1] there was nothing to be gained by the consideration of randomized denoisers (the optimum distribution-dependent scheme is always non-randomized and the universal scheme of [1] was also non-randomized), for the channel uncertainty setting of our present work randomized schemes play a key role and, in fact, minimax optimal schemes (in senses defined below) will, in general, be randomized.

may assume. 2) (c_S, a) denotes the binary S-tuple formed by the "context" c_S with the symbol a at location $\mathbf{0}$. 3) $P_{U(S)}(U_\mathbf{0} = 1|c_S)$ denotes $\Pr\left(U_\mathbf{0} = 1|U_{S\setminus\mathbf{0}} = c_S\right)$, as induced by the distribution $P_{U(S)}$. 4) $P_{U(S)}(c_S)$ is shorthand for $\Pr\left(U_{S\setminus\mathbf{0}} = c_S\right)$, as induced by $P_{U(S)}$.

Define now further J_S and f_S^* by

$$J_S\left(P_{U(S)}, \Delta, f\right) = \max_{0\le\delta\le\Delta} G_S\left(P_{U(S)}, \delta, f\right), \quad (5)$$

$$f_S^*[P_{U(S)}, \Delta] = \arg\min_{f:\{0,1\}^S\to[0,1]} J_S\left(P_{U(S)}, \Delta, f\right), \quad (6)$$

selecting an arbitrary achiever when it is not unique. Let further $\Delta_S\left(P_{U(S)}\right)$ be defined by

$$\max\left\{0 \le \delta \le 1/2 : \exists P_{X(S)} \text{ s.t. } P_{X(S)} * \delta = P_{U(S)}\right\}, \quad (7)$$

where $P_{X(S)} * \delta$ denotes the distribution of the random field obtained when corrupting $X(S)$, which is distributed according to $P_{X(S)}$, by the BSC(δ). Finally, let $\hat{Q}_S[z^{m\times n}]$ denote the empirical distribution on binary S-tuples induced by $z^{m\times n}$, i.e., $\hat{Q}_S[z^{m\times n}](u(S)) =$

$$\frac{|\{i \in V_{m\times n} : S + i \subseteq V_{m\times n}, z(S+i) = u(S)\}|}{|\{i \in V_{m\times n} : S + i \subseteq V_{m\times n}\}|}.$$

Equipped with this notation we now turn to describing the image denoiser. Let $\hat{X}_S^{m\times n}$ denote the $m \times n$ randomized binary image denoiser defined for the i's satisfying $S + i \subseteq V_{m\times n}$ by

$$\hat{X}_S^{m\times n}(z^{m\times n})[i] = f_S^*\left[\hat{Q}_S[z^{m\times n}], \Delta_S(z^{m\times n})\right](z(S+i)), \quad (8)$$

where we write $\Delta_S(z^{m\times n})$ as shorthand for $\Delta_S\left(\hat{Q}_S[z^{m\times n}]\right)$. $\hat{X}_S^{m\times n}(z^{m\times n})[i]$ can be arbitrarily defined for i's with $S + i \not\subseteq V_{m\times n}$. For space shortage we refrain from presenting $\Delta_S(z^{m\times n})$ and $f_S^*[P_{U(S)}, \Delta]$ in closed form. Suffice it to say that both have closed form expressions that are simple to evaluate. For Δ_S it is easy to obtain the input distribution on a S-tuple giving rise to $\hat{Q}_S[z^{m\times n}]$ when corrupted by a BSC(δ), and then taking the largest δ for which the said input distribution exists. This can be encapsulated particularly compactly using matrix notation[2]. The closed form for $f_S^*[P_{U(S)}, \Delta]$ is available since J_S in (5) can be expressed in closed form and then the minimization in (6) can be reduced to a simple convex optimization problem involving $2^{|S|}$ $[0,1]$-valued variables. $\hat{Q}_S[z^{m\times n}]$, which enters the denoising decision in the right side of (8) both as the first argument of $f_S^*[\cdot,\cdot]$ and as the argument of Δ_S, is acquired before commencing with the denoising by going over the image and counting the number of occurrences of each pattern in $\{0,1\}^S$ along the noisy image.

[2]The operator taking the distribution of a random field indexed by S into the distribution of the same field corrupted by the BSC can be presented as the multiplication of the simplex vector representing the noise-free distribution by the matrix which is the $|S|$-fold tensor product of the 2×2 matrix associated with the BSC.

4. PERFORMANCE GUARANTEE

For $r \ge 0$ let $\mathcal{B}_r$ denote the l_2 ball of radius r in $\mathbb{Z}^2$, i.e., $\mathcal{B}_r = \{i \in \mathbb{Z}^2 : \|i\|_2 \le r\}$. For m, n, we let our $m \times n$ binary image denoiser be given by

$$\hat{X}_{\text{univ}}^{m\times n} = \hat{X}_{\mathcal{B}_{r(m,n)}}^{m\times n} \quad (9)$$

where the right side of (9) is specified in (8) and $r(m,n) = g(\min\{m,n\})$ for some function g.

The main theoretical result of our work, whose proof is similar to that of [3, Theorem 1], is the following:

Theorem 1 *The family of randomized binary image denoisers defined in (9) is universally minimax in the sense that for all spatially stationary and ψ-mixing random fields $P_\mathbf{Z}$, there exists an unbounded increasing sequence w_t, which is a function of the mixing coefficients, such that if $g(t) \le w_t$ then* $\mathbb{MM}(P_\mathbf{Z}) \ge$

$$\sup_{\{(P_\mathbf{X},\delta): P_\mathbf{X}*\delta = P_\mathbf{Z}, \delta\in[0,\Delta(P_\mathbf{Z})]\cap\mathbb{Q}\}} L_{\{\hat{X}_{\text{univ}}^{m\times n}\}}(P_\mathbf{X}, \delta) \quad P_\mathbf{Z}-a.s.$$

The reader is referred to [4, 5] for the definition of ψ-mixing and its extension to random fields, as well as the broad range of circumstances under which it holds (in fact, all processes used in the modelling of images are ψ-mixing). The reader is referred to [3] for the explicit dependence of w_t on the mixing coefficients for one-dimensional signals, the multidimensional case being similar. The sequence of schemes defined in the above theorem is implementable with complexity, with respect to both time and memory, linear in the size of the image [6].

5. IMPLEMENTATION AND EXPERIMENTAL RESULTS

Theorem 1 guarantees the asymptotic performance of the denoiser family $\{\hat{X}_{\text{univ}}^{m\times n}\}$. The definition of this family involved the function g, which only needed to satisfy an asymptotic growth order condition and can be chosen arbitrarily on any finite subset of its domain. Thus, in practice, for finite images, the choice of the radius of the ball is subject to heuristics and experimentation (cf. discussion in [1, Section 8]).

We have implemented the minimax binary image denoiser (MMBID) of (9) for the case $m = n = 2000$, taking $r = \sqrt{2}$ as the ball radius. More specifically, this implied that the role of S in (8) is played by the 3×3 square surrounding the origin. For the calculation of Δ_S, we let the role of S be played by $\{\mathbf{0}, (-1,0), (-1,1), (0,1)\}$. There is obviously a tradeoff between the size of the neighborhoods S and the reliability of the associated empirical statistics. The choices of S, described above, were taken for the initial experiments, detailed below, and future experimentation will involve other choices of S. In particular, we have no claim or reason to believe that our choices are optimal.

We compare the performance of the MMBID to the performance of the binary image DUDE in [1]. In order to implement DUDE in the unknown channel case we make use of a scheme suggested in [2] for estimating the channel crossover probability δ. We will denote this estimate of δ as $\hat{\delta}_S$. It is not hard to show that for every S, $\Delta_S \leq \hat{\delta}_S$ with a strict inequality for all but degenerate cases. Furthermore, it can be shown that as the image dimensions and then S become large, Δ_S converges, from above, to the true $\Delta(P_{\mathbf{Z}})$ defined in (1).

The test image for our initial experimentation is a second order Markov random field (MRF) with correlation parameter α, cf. [6] for the details of the generation of these fields. We have employed the MMBID on a test image of dimensions 2000×2000, and compared its performance with that of the binary image DUDE of [1]. The DUDE was employed twice: once using $\hat{\delta}_S$, as was suggested in [2] for the setting of channel uncertainty, and once using the true channel parameter δ. This was done for several different channel parameters δ. Table 1 details the values of $\hat{\delta}_S$ and Δ_S for each channel. Denoising results are detailed in Table 2, which lists the normalized error rate of the denoised image, relative to the original one. The results show that MMBID outperforms the DUDE using $\hat{\delta}_S$, and often does as well as the DUDE with the true channel parameter δ, which is shown to be optimal in [2, 1]. Table 3 shows a sample result of denoising a text image using MMBID. As is seen, the MMBID reduces the noise significantly, though does not quite attain the performance of the optimum channel-dependent DUDE. It is arguable, however, which of the two reconstructions is of higher visual quality.

Channel parameter δ	0.01	0.02	0.05	0.1
Δ_S	0.018	0.028	0.058	0.105
$\hat{\delta}_S$	0.061	0.072	0.105	0.16

Table 1. δ values and their estimates for noise corrupted MRF with $\alpha = 0.05$

Scheme	Channel Parameters δ			
	0.01	0.02	0.05	0.1
MMBID	0.011	0.017	0.035	0.069
DUDE($\hat{\delta}_S$)	0.018	0.023	0.046	0.07
DUDE(δ)	0.009	0.017	0.035	0.068

Table 2. Denoising results for MRF with $\alpha = 0.05$

6. CONCLUSION

We have presented an algorithm for denoising binary images corrupted by a BSC of an unknown crossover parameter. The algorithm is asymptotically optimal in a minimax sense appropriate for the case of channel uncertainty, universally for all stationary sources with a mild mixing constraint. The preliminary experimental results presented seem to indicate that the denoiser has the potential of doing well on real data.

The scheme and results can be generalized to non-binary images, as well as to data indexed by higher-dimensional index sets.

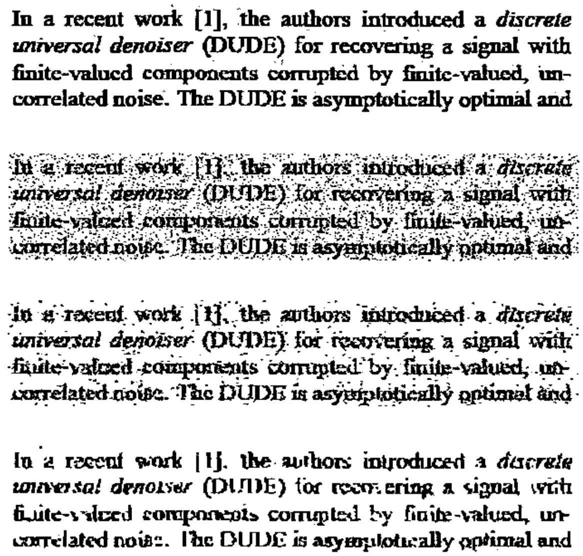

Table 3. Denoising of a 1000×600 text image. For the sake of space, we only show a 480×87 section of the images. The top image is the original image, next is the noisy version where $\delta = 0.1$, then a denoised version using MMBID with an error rate of .0479, and lastly for comparison, the DUDE(δ), i.e. with the true value, with an error rate of .0329.

7. REFERENCES

[1] E. Ordentlich, G. Seroussi, S. Verdú, M.J. Weinberger, and T. Weissman, "A universal discrete image denoiser and its application to binary images," in *Proc. IEEE International Conference on Image Processing*, Barcelona, Catalonia, Spain, Sept. 2003.

[2] T. Weissman, E. Ordentlich, G. Seroussi, S. Verdú, and M. Weinberger, "Universal discrete denoising: Known channel," 2003, To appear in IEEE Trans. Inform. Theory (preprint available).

[3] G. Gemelos, S. Sigurjònsson, and T. Weissman, "Universal discrete denoising under channel uncertainty," to appear in IEEE International Symposium on Information Theory, Chicago, Illinois, USA, June 2004.

[4] R. Bradley, "Basic properties of strong mixing conditions," in E. Eberlein and M. S. Taqqu (eds), Dependence in Probability and Statistics, Boston: Birkhuser, 1986.

[5] R. Bradley, "Equivalent mixing conditions for random fields," *Ann. Probab.*, vol. 21, pp. 1921–1926, 1993.

[6] G. Gemelos, S. Sigurjònsson, and T. Weissman, "Algorithms for universal discrete denoising under channel uncertainty," preprint available.