HIGH QUALITY IMAGE RECONSTRUCTION VIA NON-LOCAL COLLABORATIVE ESTIMATION FOR WIRELESS IMAGE/VIDEO SOFTCAST

Ruiqin Xiong[1], Jian Zhang[2], Feng Wu[3], Wen Gao[1]

[1] Institute of Digital Media, Peking University, Beijing 100871, China
[2] Department of Computer Science, Harbin Institute of Technology, Harbin 150001, China
[3] Microsoft Research Asia, Beijing 100080, China

ABSTRACT

For wireless scenarios where the channel condition fluctuates unpredictably, a novel image/video communication scheme, named SoftCast, was recently proposed to provide graceful quality degradation and competitive performance simultaneously. Unlike conventional approaches, SoftCast decorrelates input images by a transform and modulates the coefficients directly to a dense constellation for transmission, leaving out the conventional quantization, entropy coding and channel coding. The transmission is lossy in nature, with its noise level commensurate with the channel condition. To reconstruct images from the received noisy data, SoftCast employs a linear least-square estimator (LLSE), but it tends to produce annoying reconstruction artifacts. This paper proposes a high-quality image reconstruction algorithm for SoftCast, employing a collaborative estimator to utilize both the local correlation and non-local similarity within images. Experimental results show that the proposed method outperforms the existing SoftCast scheme, achieving remarkable improvement in the objective and subjective qualities of the reconstruction images.

Index Terms— Wireless video, SoftCast, image reconstruction, non-local estimation, image prior model

1. INTRODUCTION

Traditional communication systems generally require the channel condition to be known at the time of encoding, in order to choose an appropriate coding rate. Once a signal is coded and sent out, the decoding process tends to break down if the actual channel quality falls below a threshold; if the channel quality increases beyond that threshold, on the other hand, such system cannot provide any further improvement in the quality of received signal. This "threshold effect" brings great challenges for the design of wireless and mobile visual communication systems.

Recently, a scheme named *SoftCast* [1–3] was proposed for wireless video. Unlike typical image and video coders that compress input signal into a binary stream, SoftCast transforms the image signal into a stream of coefficient numbers from which exact reconstruction is possible, leaving out the conventional quantization and entropy coding. SoftCast also abandons the conventional channel coding. Instead, it modulates the number stream directly to a dense constellation for transmission. The most prominent advantage of SoftCast is that it provides graceful quality transition in very wide channel SNR range. In wireless broadcast scenarios, SoftCast can serve various clients of different channel conditions simultaneously, using the same transmitted signal in the air. For this reason, SoftCast has attracted much research attention in recent years [4–12].

The transmission in SoftCast is lossy in nature and the noise level in received data is commensurate with the instantaneous channel signal-to-noise ratio (SNR). To recover image from the noisy data received at decoder side, prior knowledge of the image should be exploited to reduce the influence of channel noises. Linear least-square estimator (LLSE) is employed in [1–3], but most channel noises remain in the results. To make things worse, the power allocation strategy in SoftCast strives to maximize the transmission performance measured by overall distortion, with a side effect that the reconstruction distortion is higher in low-frequency bands and lower in high-frequency bands. For this reason, the artifacts in the ultimate reconstruction images are usually very annoying, especially when the channel SNR is relatively low.

This paper addresses the image reconstruction problem for SoftCast. An intuitive solution is to apply a general purpose image denoising algorithm (e.g. [13–20]) to the output of SoftCast decoder. However, such denoising algorithm typically works well for additive white Gaussian noise (AWGN) but cannot efficiently handle the residual noise produced by SoftCast LLSE decoder. This paper proposes a new reconstruction algorithm for SoftCast, which solves the decoding and denoising problem jointly, employing a collaborative

Corresponding author: R. Xiong (rqxiong@pku.edu.cn). This work was supported in part by the National Natural Science Foundation of China (61370114, 61121002, 61073083), Beijing Natural Science Foundation (4132039, 4112026) and Research Fund for the Doctoral Program of Higher Education (20120001110090, 20100001120027).

estimator to utilize both the local correlation and non-local similarity within images. Experimental results show that the proposed method outperforms the existing SoftCast scheme, achieving remarkable improvement in the objective and subjective qualities of the reconstruction images.

The rest of the paper is organized as follows. Section 2 reviews the SoftCast scheme briefly. Section 3 describes the proposed reconstruction algorithm. Experimental results are reported in Section 4 and Section 5 concludes the paper.

2. REVIEW OF SOFTCAST

2.1. SoftCast Transmission

The framework of SoftCast is illustrated in Fig. 1. Input image is first decorrelated by a 2-D transform (e.g. DCT), producing a stream of transform coefficients. The SoftCast sender scales each coefficient individually, applies a Walsh-Hadamard transform (WHT) to whiten the stream, and modulates the resulted numbers to a dense constellation (e.g. 64k-QAM) for OFDM transmission. A pair of numbers from the stream is mapped to a point in the constellation, using the two numbers as the I- and the Q- components respectively, and transmitted by one OFDM sub-carrier. The receiver gets a noisy version of the stream due to channel noises. The scaling operation serves the purposes of power allocation and unequal protection against noises. The scaling factors are determined by a power-distortion optimization (PDO) procedure and shared between the sender and the receiver via a limited number of meta data. See [1–3] for more details of SoftCast.

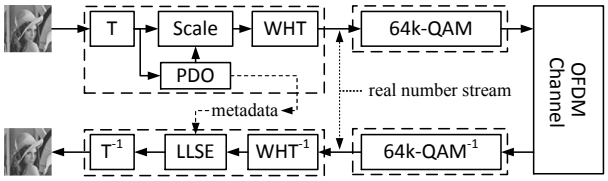

Fig. 1. Framework of the SoftCast scheme [8].

2.2. Power-Distortion Optimization

Suppose $\mathbf{x} = (x_1, x_2, \ldots, x_N) \in \mathbb{R}^N$ are the coefficients to transmit. To achieve efficient power usage, the encoder scales each coefficient x_i by a factor g_i and sends out $y_i = g_i \cdot x_i$ directly using raw OFDM (WHT is ignored in this paper[1]). After demodulation, the receiver gets $\hat{y}_i = y_i + n_i$, where n_i is channel noise. The decoder gets an estimation of x_i by $\hat{x}_i = \hat{y}_i / g_i$.[2] In this process, the expected distortion in $\hat{x}_i$ is $D_i = \mathrm{E}[(\hat{x}_i - x_i)^2] = \sigma_n^2 / g_i^2$. The transmission power for sending x_i is $P_i = \mathrm{E}[y_i^2] = g_i^2 \cdot \mathrm{E}[x_i^2]$. Therefore, the distortion-power relationship is $D_i \cdot P_i = \sigma_n^2 \cdot E[x_i^2]$.

To achieve optimal performance, the transmission power is allocated among the coefficients $\{x_i\}$ by minimizing the overall distortion $\sum_i D_i$ subject to the total power constraint $\sum_i P_i \leqslant P_{\text{total}}$. The can be easily solved by setting $\partial D_i / \partial P_i$ of all i to be equal. This eventually leads to (see [7]):

$$g_i \propto \left(E[x_i^2]\right)^{-1/4} \tag{1}$$

and $P_i = c\sqrt{\mathrm{E}[x_i^2]}$, $D_i = \frac{1}{c}\sigma_n^2\sqrt{\mathrm{E}[x_i^2]}$, where the constant c is determined by the total power P_{total}.

2.3. Distortion Characteristics of SoftCast

It is well recognized that natural images typically contain higher energy in its low-frequency bands (L-bands) but lower energy in its high-frequency bands (H-bands). For this reason, the above power allocation strategy brings two side effects in the transmission signals and reconstruction errors. Firstly, the g_i in (1) is smaller for L-bands and larger for H-bands. In effect, the power allocation procedure scales down the L-bands and scales up the H-bands, relatively. Secondly, the D_i is larger for L-bands and smaller for H-bands. That means the reconstruction error has strong low-frequency components and is not white.

To provide intuitive insight, Fig. 2 illustrates the images at various stages of the SoftCast transmission procedure. Fig. 2(a) is the original image to deliver. Fig. 2(b) is the result of power allocation, also the input to the OFDM channel (the data is actually transmitted as a stream of transform coefficients but we show it in image domain). We see that Fig. 2(b) looks like Fig. 2(a), but its low-frequency components are attenuated. Fig. 2(c) is the output from channel, containing AWGN noises. Fig. 2(d) and 2(e) are the reconstruction image and error by SoftCast LLSE decoder, respectively. It is clear that the error is annoying and not white.

3. THE PROPOSED RECONSTRUCTION SCHEME

The residual error produced by SoftCast decoder cannot be handled efficiently by postprocessing using a general purpose image denoising algorithm. In this part, we propose a scheme to solve the decoding and denoising problem jointly.

Suppose $\mathbf{u} \in \mathbb{R}^N$ is the original image, with the pixels arranged in lexicographic order, $\mathbf{x} = T\mathbf{u}$ and $\mathbf{y} = \Lambda_g \mathbf{x}$ are the results of transform and power allocation, respectively. Here T is the transform matrix and $\Lambda_g = \mathrm{diag}(g_1, g_2, ..., g_N)$ is the power allocation matrix. The transmission over noisy OFDM channel can be formulated as

$$\hat{\mathbf{y}} = \mathbf{y} + \mathbf{n}. \tag{2}$$

The reconstruction scheme aims to find a proper estimate for $\mathbf{u}$, such that $\Lambda_g T \mathbf{u}$ is close to $\hat{\mathbf{y}}$. This is an inverse problem.

[1]The Walsh-Hadamard transform for signal whitening can be ignored during power-distortion analysis, because the forward or inverse WHT transform of a white noise is still a white noise.

[2]The receiver may actually employ a linear least-square estimator (LLSE) to derive $\hat{x}_i$, if σ_n is known [1–3]. However, this aspect is ignored by the power allocation procedure since σ_n is unknown by the sender.

(a) (b) (c) (d) (e)

Fig. 2. An example of SoftCast transmission. (a) Original image, (b) Input to OFDM channel ($\{y_i\}$ shown in image domain), (c) Output from channel ($\{\hat{y}_i\}$ shown in image domain), (d) Image decoded by LLSE, (e) Reconstruction error in (d).

3.1. The Choice of Image Prior Model

To obtain stable estimation and reduce the influence of channel noise, prior knowledge of the image should be incorporated into the reconstruction scheme. According to the Bayesian rule, the *maximum a posteriori probability* (MAP) estimation of $\mathbf{u}$ can be formulated as

$$\min_{\mathbf{u}} \quad \frac{1}{2\sigma_n^2} \left\|\Lambda_g T\mathbf{u} - \hat{\mathbf{y}}\right\|_2^2 + \Phi(\mathbf{u}) \tag{3}$$

where $-\Phi(\mathbf{u}) = \log \Pr(\mathbf{u})$ is the log prior probability of $\mathbf{u}$.

Various image prior models have been proposed in literatures. Some approaches assume that patches in the image are sparse in transform domain, formulated by $\Phi(\mathbf{u}) = \sum_k \left\|\Psi\mathbf{u}_k\right\|_p$. Here $\mathbf{u}_k$ is image patch located at position k, Ψ is decorrelation transform. Some other approaches assume local smoothness in the image. For example, total variation (TV) model assumes image gradients to be close to zero, formulated by $\Phi(\mathbf{u}) = \sum_i \left\|D_i\mathbf{u}\right\|_p$ where D_i is gradient operator at position i.

To utilize both the local correlation and non-local similarity, we employ an adaptive non-local collaborative sparsity model. For each patch $\mathbf{u}_k$ in the image, we search for a group of similar patches $S_k = \{\mathbf{u}_k^1, \mathbf{u}_k^1, ..., \mathbf{u}_k^N\}$ and organize them into a 3-D data cube. We apply a 2-D PCA transform on each patch and apply another transform (e.g. DCT) along the third dimension. This forms an adaptive 3-D transform T_k for decorrelating S_k. The proposed model can be formulated as

$$\Phi(\mathbf{u}) = \sum_k \left\|\Lambda_k T_k(S_k)\right\|_p \tag{4}$$

where $\Lambda_k = \text{diag}(\frac{1}{\sigma_{k,1}}, \frac{1}{\sigma_{k,2}}, ..., \frac{1}{\sigma_{k,L}})$ and $\sigma_{k,i}$ is the estimated standard deviation for the i^{th} element of $T_k(S_k)$. The Λ_k is adaptively estimated, similar to BM3D [19].

3.2. Overall Optimization Algorithm

The optimization problem (3) is complicated and difficult to solve directly. However, it can be solved with the assistance of variable splitting and augmented Lagrangian methods [21–25]. By introducing an auxiliary variable $\mathbf{z} = \mathbf{u}$, problem (3) can be reformulated as

$$\min_{\mathbf{u},\mathbf{z}} \quad \frac{1}{2\sigma_n^2} \left\|\Lambda_g T\mathbf{z} - \hat{\mathbf{y}}\right\|_2^2 + \Phi(\mathbf{u}), \;\; \text{s.t. } \mathbf{z} = \mathbf{u} \tag{5}$$

The augmented Lagrangian function for (5) is

$$\begin{aligned} J(\mathbf{u},\mathbf{z}) = &\frac{1}{2\sigma_n^2} \left\|\Lambda_g T\mathbf{z} - \hat{\mathbf{y}}\right\|_2^2 + \Phi(\mathbf{u}) \\ &+ \frac{\beta}{2} \left\|\mathbf{z} - \mathbf{u}\right\|_2^2 + \mathbf{v}^{\mathrm{T}}(\mathbf{z} - \mathbf{u}) \end{aligned} \tag{6}$$

Here the Lagrangian variable $\mathbf{v}$ has the same dimension as $\mathbf{u}$ and $\mathbf{z}$. The problem (5) can be solved by an iterative algorithm, minimizing (6) with respect to $\mathbf{u}$ and $\mathbf{z}$ and then updating $\mathbf{v}^{\mathrm{T}}$ by $\mathbf{v}^{\mathrm{T}} \leftarrow \mathbf{v}^{\mathrm{T}} - \beta(\mathbf{z} - \mathbf{u})$ in each iteration. The minimization of (6) can be easily handled by solving the following two sub-problems.

3.3. The u Sub-problem

With $\mathbf{z}$ fixed, the problem (6) is reduced to

$$\min_{\mathbf{u}} \quad \Phi(\mathbf{u}) + \frac{\beta}{2} \left\|\mathbf{z} - \mathbf{u} - \mathbf{v}/\beta\right\|_2^2 \tag{7}$$

This is exactly the classical MAP formulation of denoising problem with AWGN noise. The solution is easily obtained by applying the image denoising algorithm (using the regularizer $\Phi(\mathbf{u})$) to the virtual noisy image $\mathbf{u}' \triangleq \mathbf{z} - \mathbf{v}/\beta$, with the standard deviation of noise on $\mathbf{u}'$ set to $\sigma = \sqrt{1/\beta}$.

3.4. The z Sub-problem

With $\mathbf{u}$ fixed, the problem (6) is reduced to

$$\min_{\mathbf{z}} \quad \frac{1}{2\sigma_n^2} \left\|\Lambda_g T\mathbf{z} - \hat{\mathbf{y}}\right\|_2^2 + \frac{\beta}{2} \left\|\mathbf{z} - \mathbf{u} - \mathbf{v}/\beta\right\|_2^2 \tag{8}$$

Since T is orthogonal, by introducing $\mathbf{z}' \triangleq \mathbf{u} + \mathbf{v}/\beta$, $\mathbf{w} \triangleq T\mathbf{z}$ and $\mathbf{w}' \triangleq T\mathbf{z}'$, the solution to (8) can be obtained by

$$\mathbf{w} = \left(\Lambda_g^{\mathrm{T}}\Lambda_g + \beta\sigma_n^2 I\right)^{-1} \cdot \left(\Lambda_g^{\mathrm{T}}\hat{\mathbf{y}} + \beta\sigma_n^2\mathbf{w}'\right) \tag{9}$$

and $\mathbf{z} = T^{-1}\mathbf{w}$. The (9) can be calculated efficiently in a component-by-component manner.

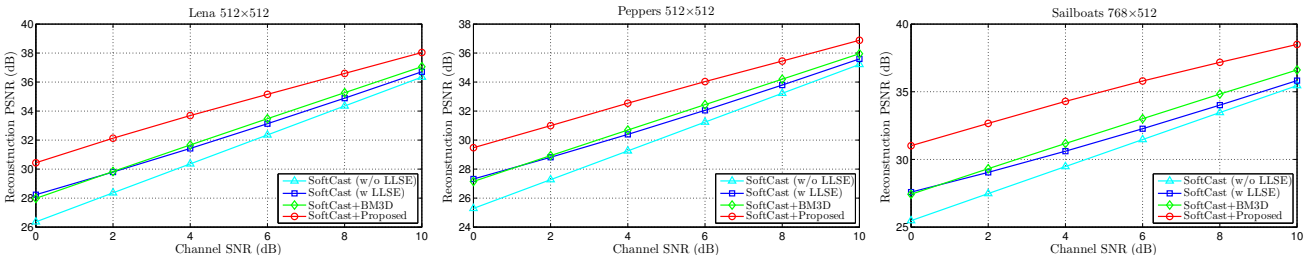

Fig. 3. PSNR results comparison between the four SoftCast reconstruction schemes.

Fig. 4. Comparison of reconstruction images produced by different SoftCast reconstruction schemes (CSNR=2dB). From left to right: (a) Original image, (b) SoftCast+LLSE, (c) SoftCast+BM3D, (d) the proposed scheme. Enlarge the figure for details.

4. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of four schemes. The first two schemes, *SoftCast (w/o LLSE)* and *SoftCast (w LLSE)*, are the original SoftCast decoder with LLSE turned off or on, respectively. The third scheme, *SoftCast+BM3D*, is the cascade of the original SoftCast with BM3D [19], a state-of-the-art image denoising algorithm. We assume the standard deviation of residual error is known by the BM3D algorithm. The fourth scheme is our proposed method described in Section 3. The method in [8] is used for power allocation in all these schemes. A large test images set is used, including [26] and [27]. The channel SNR range is set to $0 \sim 10$dB.

Fig. 3 summarized the PSNR results for *Lena*, *Peppers* and *Sailboat*. We see that *SoftCast(w LLSE)* has a gain of $0 \sim 2$dB (high at low CSNR but low at high CSNR) over *SoftCast (w/o LLSE)*. We also see that BM3D as a postprocessing only provides very slight gain. The proposed scheme outperforms other schemes by up to 3dB at low CSNR and up to 2dB at high CSNR. The improvement in subjective quality is illustrated in Fig. 4. It is clear that the proposed scheme produces much clear reconstruction images.

5. CONCLUSIONS AND DISCUSSIONS

This paper addresses the image reconstruction problem for SoftCast. Due to the power allocation strategy in SoftCast, reconstruction error produced by LLSE decoder is not white noise and cannot be easily removed by existing general purpose denoising algorithms. We propose an algorithm to solve the decoding and denoising problem of SoftCast jointly, utilizing the local correlation and non-local similarity within images simultaneously. Experimental results show that the proposed method outperforms the original SoftCast scheme, achieving remarkable improvement in the objective and subjective qualities of the reconstruction images.

6. REFERENCES

[1] Szymon Jakubczak, Hariharan Rahul, and Dina Katabi, "Softcast: One video to serve all wireless receivers," in *MIT Technical Report, MIT-CSAIL-TR-2009-005*, 2009.

[2] Szymon Jakubczak and Dina Katabi, "A cross-layer design for scalable mobile video," in *International conference on Mobile computing and networking (MobiCom '11)*, New York, NY, USA, 2011, pp. 289–300.

[3] Szymon Jakubczak, Hariharan Rahul, and Dina Katabi, "One-size-fits-all wireless video," in *The eigth ACM SIGCOMM HotNets Workshop*, New York, NY, USA, 2009.

[4] Xiaopeng Fan, Feng Wu, Debin Zhao, C. Oscar Au, and Wen Gao, "Disributed soft video broadcast (dcast) with explicit motion," in *IEEE Data Compression Conference*, 2012, pp. 199–208.

[5] Xiaolin Liu, Wenjun Hu, Qifan Pu, Feng Wu, and Yongguang Zhang, "Parcast: soft video delivery in mimo-ofdm wlans," in *Internation Conference on Mobile Computing and Networking (Mobicom)*, 2012, pp. 233–244.

[6] Xiaopeng Fan, Ruiqin Xiong, Feng Wu, and Debin Zhao, "Wavecast: Wavelet based wireless video broadcast using lossy transmission," in *IEEE Visual Communications and Image Processing (VCIP)*, 2012, pp. 1–6.

[7] Ruiqin Xiong, Feng Wu, Jizheng Xu, and Wen Gao, "Performance analysis of transform in uncoded wireless visual communication," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, Beijing, China, May 2013, pp. 1159–1162.

[8] Ruiqin Xiong, Feng Wu, Xiaopeng Fan, Chong Luo, Siwei Ma, and Wen Gao, "Power-distortion optimization for wireless image/video softcast by transform coefficients energy modeling with adaptive chunk division," in *IEEE Visual Communications and Image Processing (VCIP)*, 2013, pp. 1–6.

[9] Ruiqin Xiong, Hangfan Liu, Siwei Ma, Xiaopeng Fan, and Wen Gao, "G-cast: Gradient based image softcast for perception friendly wireless visual communication," in *IEEE Data Compression Conference (DCC)*, Snowbird, Utah, USA, Mar. 2014, pp. 133–142.

[10] Zhihai Song, Ruiqin Xiong, Xiaopeng Fan, Siwei Ma, and Wen Gao, "Transform domain energy modeling of natural images for wireless softcast optimization," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, Melbourne, Australia, June 2014, pp. 1114–1117.

[11] Zhihai Song, Ruiqin Xiong, Siwei Ma, Xiaopeng Fan, and Wen Gao, "Layered image/video softcast with hybrid digital analog transmission for robust wireless visual communication," in *IEEE International Conference on Multimedia and Expo (ICME)*, Chengdu, China, Jul. 2014.

[12] Hangfan Liu, Ruiqin Xiong, Siwei Ma, Xiaopeng Fan, and Wen Gao, "Gradient based image transmission and reconstruction using non-local gradient sparsity regularization," in *IEEE International Conference on Multimedia and Expo (ICME)*, Chengdu, China, Jul. 2014.

[13] D.L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[14] L. Yaroslavsky, K. Egiazarian, and J. Astola, "Transform domain image restoration methods: review, comparison, and interpretation," *Proc. SPIE 4304, Nonlinear Image Processing and Pattern Analysis XII*, pp. 155–169, 2001.

[15] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," *Sixth International Conference on Computer Vision (ICCV'98)*, pp. 839–846, 1998.

[16] J. Portilla, V. Strela, M.J. Wainwright, and E.P. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, 2003.

[17] A. Buades, B. Coll, and J. M Morel, "A non-local algorithm for image denoising," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 60–65, 2005.

[18] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[19] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[20] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 349–366, 2007.

[21] Wotao Yin, Stanley Osher, Donald Goldfarb, and Jerome Darbon, "Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing," *SIAM Journal on Imaging Sciences*, vol. 1, no. 1, pp. 143–168, Mar. 2008.

[22] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.

[23] T. Goldstein and S. Osher, "The split bregman method for l1-regularized problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, 2009.

[24] M.V. Afonso, J.M. Bioucas-Dias, and M. A T Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2345–2356, Sept 2010.

[25] M.V. Afonso, J.M. Bioucas-Dias, and M. A T Figueiredo, "An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 681–695, March 2011.

[26] "The usc-sipi image database," `http://sipi.usc.edu/database/`.

[27] "True-color kodak test images," `http://r0k.us/graphics/kodak/`.