

Sparse super-resolution reconstructions of video from mobile devices in digital TV broadcast applications - PS-2006-0117

Choong S. Boon*, Onur G. Guleryuz†, Toshiro Kawahara*, Yoshinori Suzuki*

*Research Laboratories, NTT DoCoMo, Inc.
3-5 Hikarino-oka, Yokosuka,
Kanagawa, 239-8536, Japan
{boon,kawahara,ys}@mml.yrp.nttdocomo.co.jp

† DoCoMo Communications Laboratories USA, Inc.
3240 Hillview Avenue
Palo Alto, CA 94304, USA
guleryuz@docomolabs-usa.com

ABSTRACT

We consider the mobile service scenario where video programming is broadcast to low-resolution wireless terminals. In such a scenario, broadcasters utilize simultaneous data services and bi-directional communications capabilities of the terminals in order to offer substantially enriched viewing experiences to users by allowing user participation and user tuned content. While users immediately benefit from this service when using their phones in mobile environments, the service is less appealing in stationary environments where a regular television provides competing programming at much higher display resolutions. We propose a fast super-resolution technique that allows the mobile terminals to show a much enhanced version of the broadcast video on nearby high-resolution devices, extending the appeal and usefulness of the broadcast service. The proposed single frame super-resolution algorithm uses recent sparse recovery results to provide high quality and high-resolution video reconstructions based solely on individual decoded frames provided by the low-resolution broadcast.

Keywords: mobile broadcast, terrestrial broadcast, digital tv, iterated denoising, recovery, prediction

1. INTRODUCTION

As mobile phones and other wireless devices are closing in on achieving ubiquity, they are becoming plausible destinations for the delivery of multimedia content, especially for the delivery of video. While mobile devices have limited general purpose computational and display resources, better radio technologies, increasing coverage areas, and high downlink speeds encourage their use as intelligent content delivery stations outside of the home. Interestingly with rapid increases in immersive mobile services, many users of mobile technologies would also like to have access to information and services from their mobile lives in stationary environments.



Figure 1. NTT DoCoMo FOMA P901iTV: An advanced mobile terminal capable of receiving mobile digital TV broadcasts. The phone is equipped with a special antenna and other advanced functionality such as video recording, programming/Internet integration using established mobile web interfaces like i-mode, etc.^{22, 23}

In this paper we are primarily interested in a scenario where video and/or TV programming is broadcast to mobile phones.²² Such broadcasting is currently available in Japan and is in the process of being deployed in the US and in other areas of the world. While users immediately benefit from this service when using their phones in mobile environments, the service loses its appeal in stationary environments where a regular television provides competing programming at much higher display resolutions and quality. As the broadcast service is intended for the mobile device and its display, broadcasting high resolution video solely for a secondary application in a stationary environment is very wasteful. The low resolution and quality of the broadcast video thus becomes an important practical bottleneck for the full adoption of such interesting wireless video broadcast services.

In order to overcome this bottleneck, we propose an algorithm that provides super-resolution reconstructions of low resolution video received by mobile phones equipped with technologies that allow reception of broadcast video. Our algorithm super-resolves video decoded by the wireless device for display on a nearby high resolution display device, such as a television. As high quality multiframe super-resolution requires optical flow caliber motion estimation, we restrict ourselves to operate on single frames and in some sense solve an interpolation problem*. The resulting algorithm can be deployed on the phone, or if resources do not avail, on a separate interface to the high resolution display device that implements the required computation. The single frame mode does not introduce delay and serves to limit the need for buffers and expensive memory operations. In case of high-end resources, our formulation can also be easily extended to use multiple frames.

The algorithm we propose is based on recent sparse recovery results and provides high quality especially around edges and other image singularities.¹³ Using an overcomplete basis that is expected to provide sparse descriptions over video frames, our algorithm recovers high resolution data by assuming that the decoded low resolution video frame forms the low frequency band of a two dimensional wavelet transform. The problem thus becomes the estimation of high frequency wavelet coefficients based on the knowledge of the low frequency wavelet band. Overcoming the need for an ill-conditioned numerical inverse, the bulk of the algorithm thus serves to enforce the locally sparse statistical model in order to estimate the missing high frequency data. The model is formed via the utilized basis with the aid of weighted overcomplete denoising.¹⁵ Our work involves fast computations that implement the main super-resolution technique followed by well established enhancement and color correction techniques appropriate for final display.

Sections 2 provides a brief overview of terrestrial digital TV broadcasts to mobile terminals in Japan followed by Section 3 which sets up our super-resolution model specialized to this application. Quality improvements that can be expected and some limitations are detailed in Section 3.1 with illustrative examples from the proposed work. Section 4 provides our basic formulation and Section 5 incorporates simulation results. We conclude the paper in Section 6.

2. OVERVIEW OF TERRESTRIAL DIGITAL TV BROADCASTS TO MOBILE TERMINALS

In this paper we are primarily interested terrestrial digital TV broadcasts to mobile devices such as the services that recently started in Japan. Following the work in,²² this section presents a brief technical overview.

The broadcasts are in the UHF band, and depending on the utilized modulation technique, involve 312, 416, or 624 kilobits per second data rates per channel. The transmitted video is compressed using h.264/AVC¹⁶ and audio compression is achieved via MPEG-2 AAC.⁴ Video is transmitted at 15 frames per second. Reception is via mobile phones with advanced capabilities including a recording functionality (Figure 1).

Using simultaneous data transmission and broadcaster specific storage areas on the phones, users can be involved in an immersive viewing experience that allows them to participate on quiz shows, vote on programming, etc. Users can also directly link to Internet sites related to the programming via programming specific web content that is matched to wireless devices using technologies such as i-mode.²²

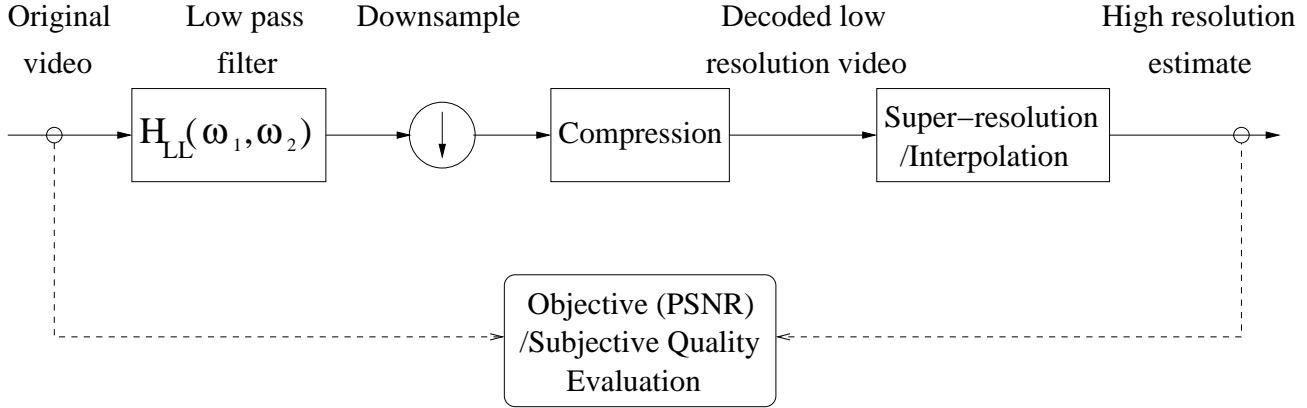


Figure 2. The super-resolution setup considered in this paper. Original video frames are spatially low pass filtered and downsampled. The low-resolution frames undergo compression and decompressed low-resolution video frames are used by the super-resolution algorithm to form a high-resolution estimate.

3. SUPER-RESOLUTION MODEL

Figure 2 illustrates the model considered in this paper. High-resolution video frames are spatially low-pass filtered with a filter with frequency response $H_{LL}(\omega_1, \omega_2)$. The filtered frames are then downsampled (without loss of generality, we will assume downsampling by two in each spatial dimension) and the downsampled video frames are compressed with a video codec (we will assume compression is via h.264/AVC). At the decoding end the video is decompressed and fed into a super-resolution algorithm which is responsible for obtaining a good approximation to the original high-resolution video. Evaluation is done with respect to comparing the reconstructed high-resolution video with the original video. The reader should note that this is a simplified model compared to more general super-resolution setups, where different frames may undergo different filtering, spatial warping, etc..^{1, 8, 9, 19}

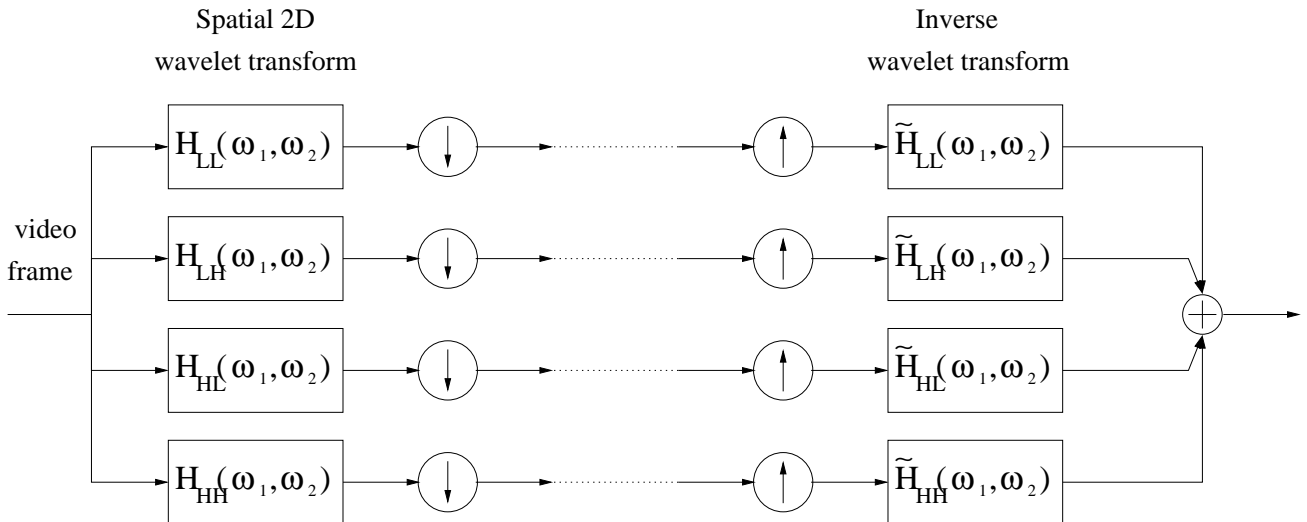


Figure 3. In this paper we assume that $H_{LL}(\omega_1, \omega_2)$ can be complemented to a perfect reconstruction wavelet filter bank so that the input and reconstructed video frames are identical.

In order to avoid ill-conditioned inverse problems, we will assume that the filter $H_{LL}(\omega_1, \omega_2)$ can be comple-

*Single frame super-resolution is often times considered to be an interpolation problem. However, since our technique effectively predicts missing high frequency wavelet coefficients (see Section 3), it is also solving an extrapolation problem.

mented to a perfect reconstruction, biorthogonal wavelet filter bank,¹³ i.e., $H_{LL}(\omega_1, \omega_2)$ is such that it is part of the perfect reconstruction filter-bank shown in Figure 3.²⁰ We note however that our work can be generalized in a straightforward way to cases where $H_{LL}(\omega_1, \omega_2)$ does not admit a compact, perfect reconstruction dual. Using this filter-bank setup, our model is as represented in Figure 4. Note that estimation of high-resolution video frames boils down to the estimation of the high frequency subband coefficients in each video frame based on the available low frequency subband, i.e., one can consider the missing data recovery problem where the high frequency subbands are lost and the low frequency subband is used to estimate the missing data.

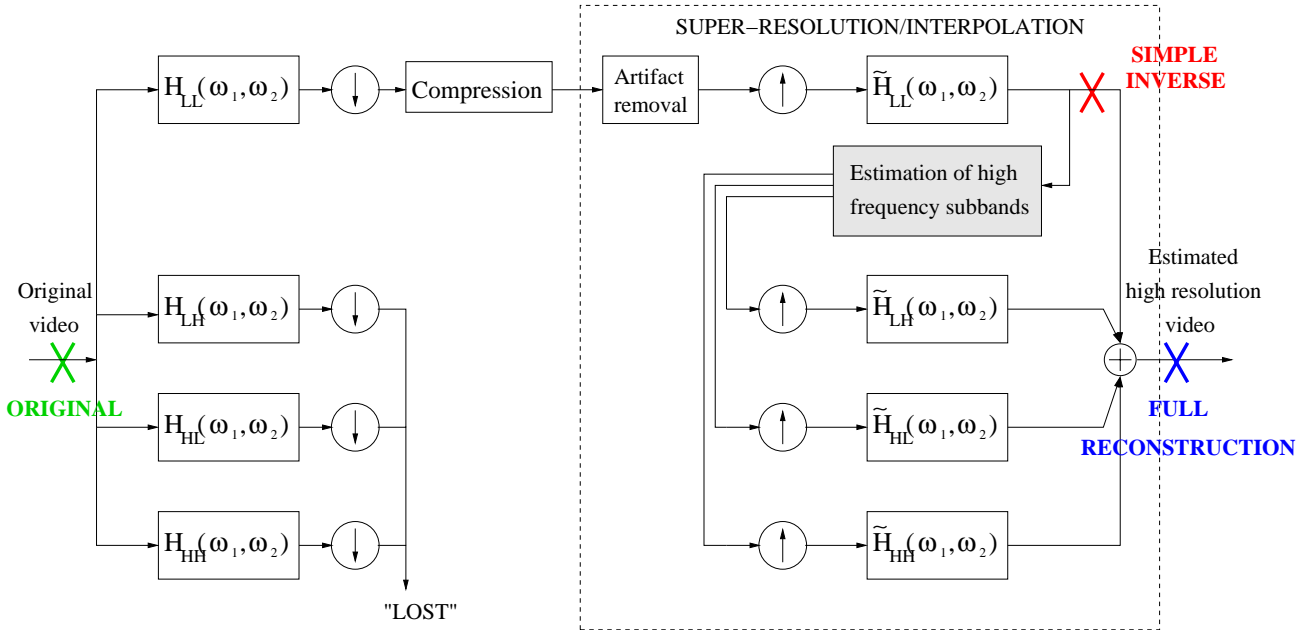


Figure 4. Single frame super-resolution as estimation of “lost” high-frequency subbands of a wavelet transform. The decompressed low-resolution video is passed through an algorithm that removes compression artifacts. Using the result as the low-resolution subband of a wavelet transform and inverse transforming results in the “Simple Inverse”. The nonlinear algorithm uses this “Simple Inverse” nonlinearly to obtain estimates of lost high-frequency subbands. The inverse transform after the estimation results in the full reconstruction. Since the low-frequency subband remains unchanged, the full reconstruction always agrees with the available data, i.e., low-pass filtering it with $H_{LL}(\omega_1, \omega_2)$ and downsampling results in the artifact removed, decompressed low-resolution video.

While one can use other available low-resolution frames in addition to the current frame to help with estimation, for reasons of complexity, delay, and memory size limitations, we will concentrate on estimation using the current frame only. Our solutions thus form single frame super-resolution reconstructions and in some ways our setup is similar to the scenarios considered in.^{5,6,17} Regardless of the limitation to a single frame, as we will see, we obtain high quality reconstructions both in objective and subjective evaluations especially around edges and other singularities in the video frames. We also note that our techniques can easily be extended to utilize multiple frames at the expense of complexity and delay.

We are interested in two types of evaluations of the generated high-resolution approximations. In cases where $H_{LL}(\omega_1, \omega_2)$ is known we will use objective reconstruction quality (as determined via PSNR measurements) in addition to subjective quality. When $H_{LL}(\omega_1, \omega_2)$ is unknown, we will turn to subjective evaluations only. In our evaluations we will consider the “simple inverse”, which is the output obtained by linear interpolation of the low-resolution video, and the “full reconstruction”, which is the full nonlinearly reconstructed video (Figure 4). Observe that our setup guarantees that the full reconstruction will agree with the available data, i.e., when the full reconstruction is low pass filtered with $H_{LL}(\omega_1, \omega_2)$ and downsampled, we obtain the decompressed low-resolution video after compression artifact removal.

3.1. Expected Improvements and Limitations

It is clear that the estimation procedure shown in Figure 4 can guess the high-frequency subbands accurately (in the mean squared error sense) only when the high frequency subbands are predictable from the low frequency subband. It is hence worthwhile to briefly talk about stochastic processes that model images and whether accurate prediction over subbands is possible using these models. While we haven't provided the details of our technique yet, we provide the discussion in light of simulation results obtained by running our method on uncompressed images and video in order to provide concrete examples. For the simulation results in this section, we assume the filter-bank in our model is formed using Daubechies $D7 - 9$ wavelets.

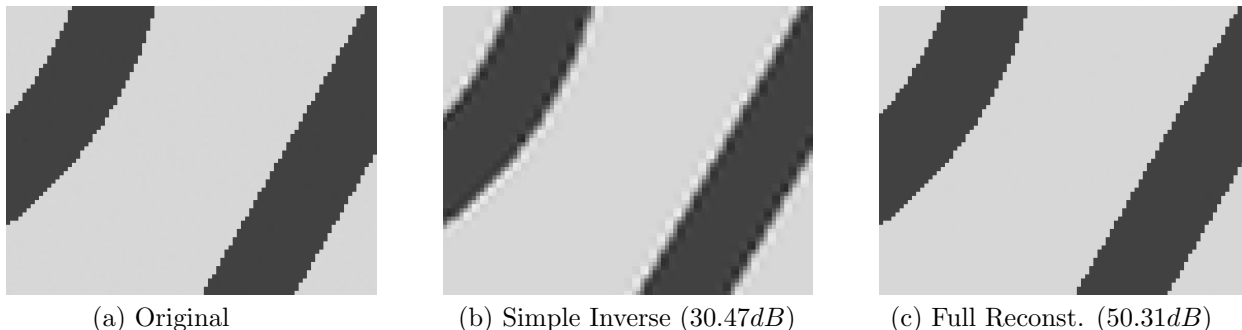


Figure 5. A piecewise smooth image with sharp edges. The baseline simple inverse (refer to Figure 4 for definitions) is able to reconstruct a blurry image with artifacts around edges but is able to recover the smooth regions well. The full reconstruction improves significantly over this baseline ($\sim +20dB$) by predicting the high frequency subbands accurately. In this example the high frequency subbands are non-zero only over and around the edges, precisely the information lacking in the simple inverse. The full reconstruction obtains all improvements by predicting the wavelet coefficients on or around edges.^{13, 14}

Piecewise Smooth Processes (Figure 5): Images are typically modeled as piecewise smooth processes. Over such processes it is well-known that good filter banks accomplish decorrelation over smooth parts of the process, i.e., on smooth portions of the process accurate prediction across subbands is not possible. Yet it is also well known that such filter banks have excellent energy compaction properties over the smooth portions, to the degree that the low resolution subband approximately contains all of the information about the smooth portions, i.e., on smooth portions of the process accurate prediction is mostly inconsequential. On the other hand, researchers have extensively noted that over edges/singularities of piecewise smooth processes, subband coefficients have strong statistical dependencies (see for example^{2, 7}) and as we have shown, accurate prediction of high-frequency subbands is possible and leads to significant improvements.¹³

Piecewise Uniform Processes (Figure 6): A more sophisticated model is one where images are considered to be decomposed of regions of uniform statistics (smooth, high-frequency, texture, etc.). Using the intuition provided by successful texture models defined in wavelet domain,²¹ we can again say that on uniform portions of the image accurate prediction is difficult since good filter banks will generally lead to decorrelated subbands. Unlike the case for piecewise smooth processes however, if the uniform regions contain high frequencies, the low resolution subband will not contain all the information needed to reconstruct the region. Hence the lack of accurate prediction over uniform high-frequency and texture regions is expected to dominate the mean squared error. Note that, with the techniques outlined in this work, it is still possible to obtain high accuracy predictions around edges and singularities. But due to high mean squared error obtained over texture and high-frequency regions that may be present, accurate prediction leads to less significant overall improvements in PSNR.^{13, 14}

These simple models can only partially govern real world images and video, which may manifest many different types of sophisticated variations. Regardless, the intuition provided by the simple models remains to be approximately valid. This is illustrated in Figure 7, where the full reconstruction obtains significant subjective and objective improvements around frame singularities. For detailed analysis about expected improvements in more general setups we refer the reader to.^{3, 18}

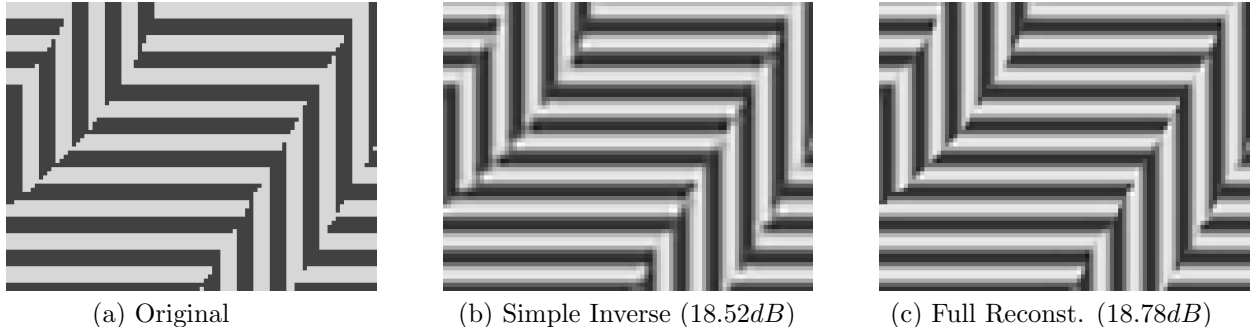


Figure 6. The same singularity geometry as in Figure 5 but smooth regions replaced with higher frequency components. The simple inverse is able to reconstruct an image with artifacts around edges. Furthermore high frequency portions have been rounded off to sinusoids since the frequency components that add to form the square waves seen in the original have been eliminated by $H_{LL}(\omega_1, \omega_2)$. The full reconstruction obtains significant improvements around the edges but it is not able to improve over the uniform regions. While subjective improvements are good, objective improvements ($\sim +0.25dB$) are marginal since the bulk of the mean squared error is due to lost high frequency components in uniform regions. When one evaluates PSNR inside a thin shell around the edges, objective improvements jump to $\sim +1.5dB$, indicating that the full reconstruction is still able to guess the coefficients around and on edges with some accuracy.

4. BASIC FORMULATION

Different from the wavelet transform used in our model in Figure 4, our formulation uses a set of sparsity enforcing transforms that allow the recovery of the missing subbands. For fast computations and for their ability in providing sparse decompositions over locally uniform regions, we use a block ($B \times B$) DCT and its B^2 shifts as the sparsity enforcing transforms so that translation invariant operation is ensured.

Suppose the original frame is lexicographically ordered into a $(N \times 1)$ vector x and assume that we are given a linear orthonormal transform \mathbf{G} ($N \times N$). Let g_i^T , the i^{th} row of \mathbf{G} , denote the i^{th} transform basis function and let $c_i = g_i^T x$ be the corresponding transform coefficient. We have

$$x = \sum_{i=1}^N c_i g_i. \quad (1)$$

We assume that \mathbf{G} generates a sparse decomposition of x so that most of the transform coefficients of x are zero or close to zero. Define the significant set $\mathcal{S}(x, K)$ as the indices of the K largest in magnitude coefficients of x with $K \ll N$. Then our assumption can be stated as

$$x = \sum_{i \in \mathcal{S}(x, K)} c_i g_i + \sum_{i \notin \mathcal{S}(x, K)} c_i g_i \cong \sum_{i \in \mathcal{S}(x, K)} c_i g_i, \quad (2)$$

that is nonlinear approximation of x using transform \mathbf{G} with the largest K coefficients closely approximates x . Since the transform is orthonormal, (2) amounts to assuming that $|c_i| \cong 0$, $i \notin \mathcal{S}(x, K)$.

Our formulation starts with an approximation to the original frame, y_0 . We take this approximation as the simple inverse shown in Figure 4. Based on this approximation we obtain a significant set $\mathcal{S}(y_0, K_0)$ which allows us to establish sparsity constraints of the form $|c_i| \cong 0$, $i \notin \mathcal{S}(y_0, K_0)$. Solving these constraints jointly with the available information given by the values of the low frequency subband coefficients (which in turn determine linear constraints), allows us to obtain the next approximation y_1 . We then repeat the process by obtaining a new significant set $\mathcal{S}(y_1, K_1)$, establish new sparsity constraints, solve jointly with available information, and so on. In practice using a single sparsity enforcing transform is not sufficient so we use an overcomplete set of orthonormal transforms \mathbf{G}_j , $j = 1, \dots, M = B^2$ as given by the shifts of the DCT. The sparsity constraints are then determined via least squares, and the significant sets are obtained by thresholding the transform coefficients using a sequence of decreasing thresholds. It can be shown that this formulation can be written as a sequence of denoising (with hard thresholding) and available data projection operations.^{10, 11, 13} We also augment the



(a) Original



(b) Low-res.



(c) Simple Inverse (Y: 32.34dB)



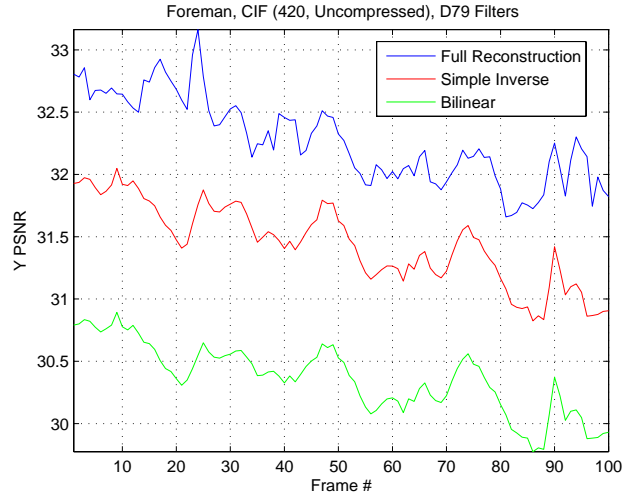
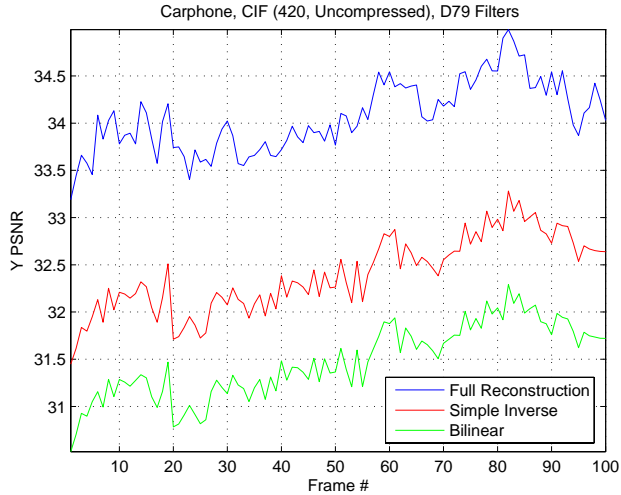
(d) Full Reconst. (Y: 33.74dB)

Figure 7. Carphone sequence, frame 183, uncompressed. (a) Original at CIF (420) resolution, (b) Low resolution (QCIF) shown twice the size using pixel replication, (c,d) Simple Inverse and Full Reconstruction. The full reconstruction obtains significant improvements around edges to yield a sharper image with better objective ($\sim +1.4dB$) and subjective quality. While accuracy around edges is good, some missing high frequency detail such as that over the text in the upper left corner and texture on the trees on the right cannot be recovered.

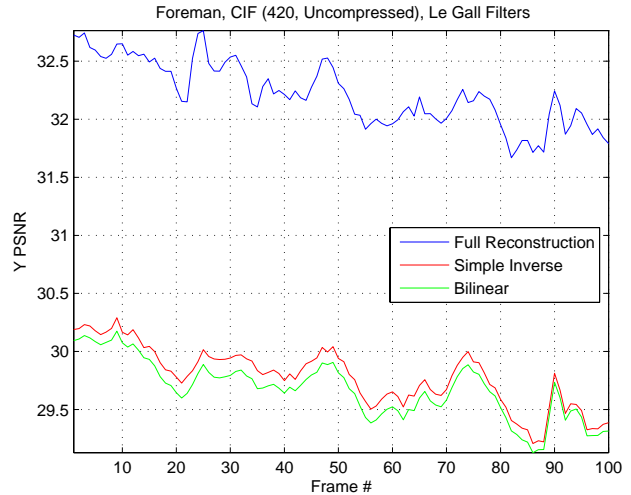
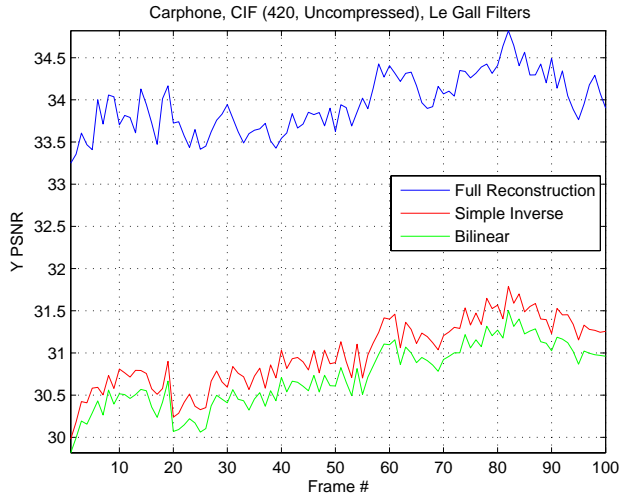
denoiser to use weighted denoising for better performance around edges. For details of the formulation we refer the reader to.^{10, 11, 13}

5. SIMULATION RESULTS

For the results in this section original frames are at CIF resolution. Compression artifact removal is via the technique outlined in.¹² We first consider the case where the low pass filter $H_{LL}(\omega_1, \omega_2)$ in Figure 2 is known and conforms to our wavelet transform model. In this case we can evaluate the objective performance of the proposed algorithm. This is done in Figure 8, where we consider two different filter banks, the $D7-9$ filters and the Le Gall filters, assuming no compression. As can be seen, the full reconstruction consistently outperforms the simple inverse significantly. For illustration purposes we also show the performance of bilinear interpolation which performs below the simple inverse. This is not surprising since using bilinear interpolation is equivalent to using a mismatched filter in place of the $\tilde{H}_{LL}(\omega_1, \omega_2)$, i.e., rather than the correct inverse transform one is evaluating a mismatched inverse transform.



(a) D79 filters (impulse response of $H_{LL}(\omega_1, \omega_2)$ is 9×9).

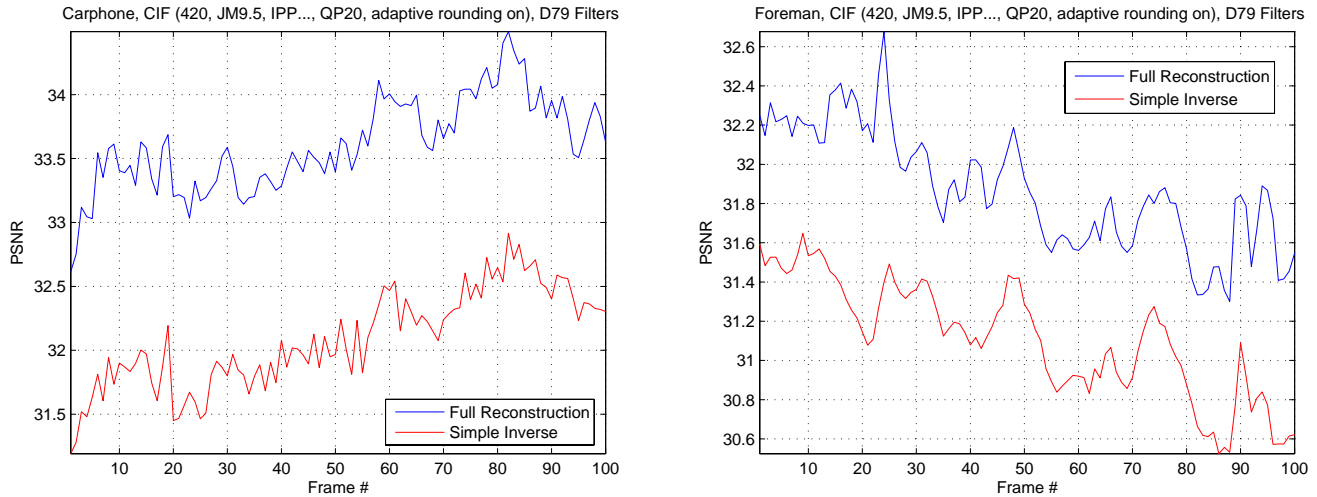


(b) Le Gall filters (impulse response of $H_{LL}(\omega_1, \omega_2)$ is 3×3).

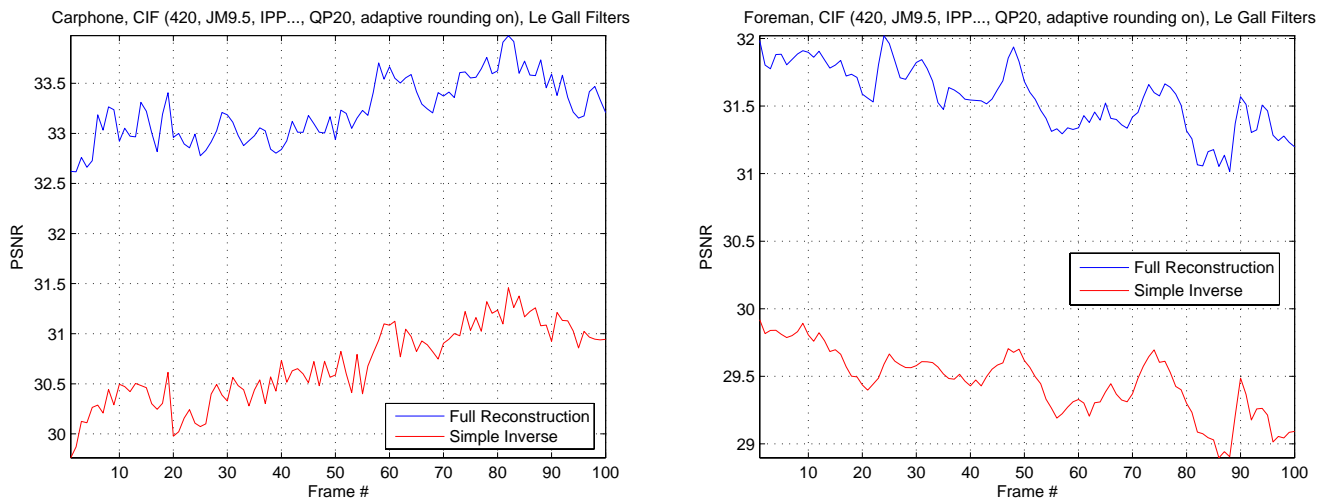
Figure 8. Objective single frame super-resolution results using known $H_{LL}(\omega_1, \omega_2)$. The proposed full reconstruction technique obtains significantly better reconstructions compared to the simple inverse. The simple inverse in turn outperforms straightforward bilinear interpolation. In (a) $H_{LL}(\omega_1, \omega_2)$ is given by the forward low-frequency filter of the D79 filter bank. In (b) $H_{LL}(\omega_1, \omega_2)$ is given by the forward low-frequency filter of the Le Gall filter bank.

In Figure 9 we show the same set of results for the case when the low resolution video has been compressed by h.264/AVC. Compression details are as outlined in the figure.

Finally in Figures 10 and 11 we provide subjective results for the case when $H_{LL}(\omega_1, \omega_2)$ is not known and does not conform to our wavelet transform model. The proposed full reconstruction again obtains subjective improvements. These improvements also become pronounced when one wishes to augment these results with color correction and sharpening for better subjective quality before final display. Since the simple inverse lacks high frequency information sharpening can only enhance intermediate frequencies, resulting in thick edges and ringing artifacts.



(a) D79 filters.



(b) Le Gall filters.

Figure 9. Objective results using the same setup as Figure 8 but with compressed input. Compression is via h.264/AVC reference codec, JM 9.5 with $QP = 20$ and IPP... encoding. Results and improvements are basically similar to Figure 8, with some performance lost due to compression.

6. CONCLUSION

As mobile devices become more prevalent entertainment platforms, their interface to high-resolution, high-capability devices in stationary environments becomes important. Such interfaces allow mobile users to have seamless and high quality continuity between their mobile lives and stationary environments. We proposed a single frame super-resolution algorithm that is targeted toward increasing the resolution of mobile broadcast video, effectively forming a high quality conduit between a mobile device and a high-resolution display for broadcast video applications. The proposed algorithm is efficient computationally and in memory usage. As our results on video sequences show, the algorithm can render high-resolution frames with good subjective quality especially around image edges and singularities. When the low-pass filter generating the low-resolution data is known, the algorithm also provides significant objective (PSNR) improvements around such regions.



(a) Original



(b) Low-res.



(c) Simple Inverse



(d) Full Reconst.



(e) Simple Inverse (Color corr. + unsharp masking)



(f) Full Reconst. (Color corr. + unsharp masking)

Figure 10. Subjective results. Carphone sequence, frame 7. (a) Original at CIF (420) resolution, (b) Low resolution (QCIF), h.264/AVC compressed at $QP = 20$. Shown twice the size using pixel replication, (c,d) Simple Inverse and Full Reconstruction. (e,f) Simple Inverse and Full Reconstruction after the application of color correction and sharpening.



(a) Original



(b) Low-res.



(c) Simple Inverse



(d) Full Reconst.



(e) Simple Inverse (Color corr. + unsharp masking)



(f) Full Reconst. (Color corr. + unsharp masking)

Figure 11. Subjective results. Foreman sequence, frame 5. (a) Original at CIF (420) resolution, (b) Low resolution (QCIF), h.264/AVC compressed at $QP = 20$. Shown twice the size using pixel replication, (c,d) Simple Inverse and Full Reconstruction. (e,f) Simple Inverse and Full Reconstruction after the application of color correction and sharpening.

REFERENCES

1. Y. Altunbasak, A. Patti, and R. Mersereau. "Super-resolution still and video reconstruction from mpeg-coded video," *IEEE Trans Circuits Systems for Video Tech.*, vol 12, 2002.
2. F. Arandiga, A. Cohen, M. Doblaz, and B. Matei, "Edge Adapted Nonlinear Multiscale Transforms for Compact Image Representation," *Proc. IEEE Int. Conf. Image Proc.*, Barcelona, Spain, 2003.
3. S. Baker and T. Kanade, "Limits on Super-Resolution and How to Break Them," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol: 24, No. 9, September 2002.
4. M. Bosi, K. Brandenburg, S. Quackenbusch, K. Akagiri, H. Fuchs, J. Herre, L. Fielder, M. Dietz, Y. Oikwa and G. Davidson, "ISO/IEC MPEG-2 Advanced Audio Coding", presented at the 101st Convention of the Audio Engineering Society, J.Audio Eng. Soc. (Abstracts), vol. 44, p. 1174, Dec. 1996.
5. W. K.Carey, D. B. Chuang, and S. S. Hemami, "Regularity-Preserving Image Interpolation," *IEEE Trans. on Image Proc.*, vol. 8, No. 9, pp. 1293-1297, 1999.
6. S. G. Chang, Z. Cvetkovic, and M. Vetterli, "Locally-Adaptive Wavelet-Based Scheme for Image Interpolation," *IEEE Transactions on Image Processing*, June 2006.
7. M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. on Signal Proc.*, vol. 46, pp. 886-902, 1998.
8. M. Elad and A. Feuer, "Super-resolution reconstruction of image sequences," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol:21, pp:817-834, 1999.
9. S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multi-frame super-resolution," *IEEE Trans. Image Processing*, October, 2004.
10. O. G. Guleryuz, "Nonlinear Approximation Based Image Recovery Using Adaptive Sparse Reconstructions and Iterated Denoising: Part I - Theory," *IEEE Trans. on Image Processing*, vol. 15, No. 3, pp. 539-554, March, 2006.
11. O. G. Guleryuz, "Nonlinear Approximation Based Image Recovery Using Adaptive Sparse Reconstructions and Iterated Denoising: Part II - Adaptive Algorithms," *IEEE Trans. on Image Processing*, vol. 15, No. 3, pp. 555-571, March, 2006.
12. O. G. Guleryuz, "A Nonlinear Loop Filter for Quantization Noise Removal in Hybrid Video Compression," *Proc. IEEE Int'l Conf. on Image Proc. (ICIP2005)*, Genova, Italy, Sept. 2005.
13. O. G. Guleryuz, "Predicting Wavelet Coefficients Over Edges Using Estimates Based on Nonlinear Approximants," *Proc. Data Compression Conference*, IEEE DCC-04, April 2004.
14. O. G. Guleryuz, "Predicting Wavelet Coefficients Over Edges Using Estimates Based on Nonlinear Approximants," DCC-04 presentation slides, April 2004. http://eeweb.poly.edu/onur/online_pub.html
15. O. G. Guleryuz, "Weighted Overcomplete Denoising," *Proc. Asilomar Conf. on Signals and Systems*, Pacific Grove, CA, Nov. 2003.
16. Joint Video Team of ITU-T and ISO/IEC JTC 1, "Draft ITU T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 — ISO/IEC 14496-10 AVC)," Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT-G050, March 2003.
17. X. Li and M. T. Orchard, "New edge-directed interpolation", *IEEE Trans. on Image Processing*, vol: 10, no. 10, October 2001.
18. D. Robinson and P. Milanfar, Statistical Performance Analysis of Super-Resolution, *IEEE Trans. on Image Processing*, vol: 15, No. 6, June 2006.
19. C.A. Segall, R. Molina, A. Katsaggelos, and J. Mateos. "Bayesian highresolution reconstruction of low-resolution compressed video," *Proc. IEEE Int'l Conf. on Image Proc. (ICIP2001)*, Oct. 2001.
20. M. Vetterli and J. Kovacevic, "Wavelets and Subband Coding," Prentice Hall, Englewood Cliffs, NJ, 1995.
21. A. S. Willsky, "Multiresolution Markov models for signal and image processing", *Proceedings of the IEEE*, vol:90, no:8, August 2002.
22. O. Yamamori, H. Atsumi, T. Mizoguchi, and K. Ishii, "Mobile Terminal Supporting Terrestrial Digital TV Broadcasting," NTT DoCoMo Technical Journal, Vol. 8, No. 1, pp. 47-56.
23. http://www.nttdocomo.co.jp/product/concept_model/p901itv/