Example-based Enhancement of Degraded Video

Edson M. Hung, *Member, IEEE*, Diogo C. Garcia, *Member, IEEE*, and Ricardo L. de Queiroz, *Senior Member, IEEE*

***Abstract*—We present an example-based approach to general enhancement of degraded video frames. The method relies on building a dictionary with non-degraded parts of the video and to use such a dictionary to enhance the degraded parts. The image degradation has to originate from a "repeatable" process, so that the dictionary image patches (blocks) are equally degraded, thus originating a dictionary with degraded blocks and their residues (differences in between degraded and original blocks). Once a match is found between a degraded block in the video and a degraded block in the dictionary, the associated residue of the latter is soft-added to the block of the former. The method is a generalization of the method for example-based super-resolution. Results are presented to demonstrate the applicability of the method to many scenarios.**

***Index Terms*—Example-based, noise removal, super-resolution, video enhancement.**

I. Introduction

WE CONSIDER an enhancement technique for degraded video that relies on examples, i.e. based on codebooks containing examples of "how non-degraded images should look like". The proposed work is a generalization of super-resolution-by-example methods [1], [2].

In the literature [3]–[5], many approaches for super-resolution can be found and are usually classified as frequency- and spatial-based-domain. In some works on frequency-domain super-resolution, the authors also extend the super-resolution problem by adding noise and blur into low-resolution images [6], [7].

A specific application of the super-resolution problem is in mixed-resolution video, i.e., in video with different resolutions along time. The solutions presented in previous works [2], [8] avoid an ill-posed problem by using key-frames as example. In those, dictionaries are constructed as examples of high-resolution images. Patches of low-resolution images are then matched to the low-resolution version of the dictionary entries. Once a match is found, the low-resolution image is super-resolved with the aid of the full-resolution entry. Such a method is here extended and adapted to general repeatable forms of image degradation.

Manuscript received January 31, 2014; revised March 25, 2014; accepted May 11, 2014. Date of publication May 22, 2014; date of current version May 29, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Zhu Liu.

The authors are with the Gama Engineering School, Universidade de Brasilia, Brasília - DF, 70910-900, Brazil (e-mail: mintsu@image.unb.br; diogo@image.unb.br; queiroz@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2014.2326551

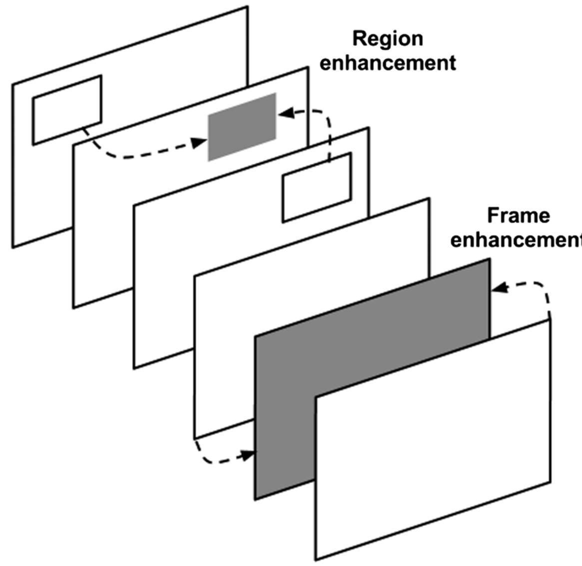

Fig. 1. A degraded frame or a region degradation are enhanced using information from neighboring frames.

Video naturally provides excellent examples to construct dictionaries because temporally adjacent frames are usually very correlated. The proposed enhancement is illustrated in Fig. 1.

There are related works based on video quality enhancement [9], spatio-temporal filtering [10], video debluring [11], or video denoising. Studies about flickering [12] also yield video enhancement based on temporal correlation. To the best of our knowledge, we are the first to use an example-based approach for video enhancement, which are suitable for cloud-based applications [13]. Details of the proposed method are presented in the next section.

Some application scenarios wherein degradation would not affect the whole video sequence include the use of multiple-frame resolution in distributed video coding, mixed-quality video coding, or situations during video streaming when the frame quality may change depending on network restrictions, fortuitous errors or autofocus delay.

II. Generalized Example-Based Enhancement

Let $\mathbf{x}, \mathbf{y} \in \Re^N$ such that

$$F(\mathbf{x}) = \mathbf{y} \tag{1}$$

where F represents a repeatable non-invertible process. By repeatable we mean that it does not depend upon parameters beyond user control (e.g. adding random noise), i.e. if $\mathbf{x}_i = \mathbf{x}_j$ then $F(\mathbf{x}_i) = F(\mathbf{x}_j)$. Since it is not invertible, there is no F' such that $F'(\mathbf{y}) = \mathbf{x}, \forall \mathbf{x} \in \Re^N$. Thus, the space of all $\{\mathbf{y}_i\}$

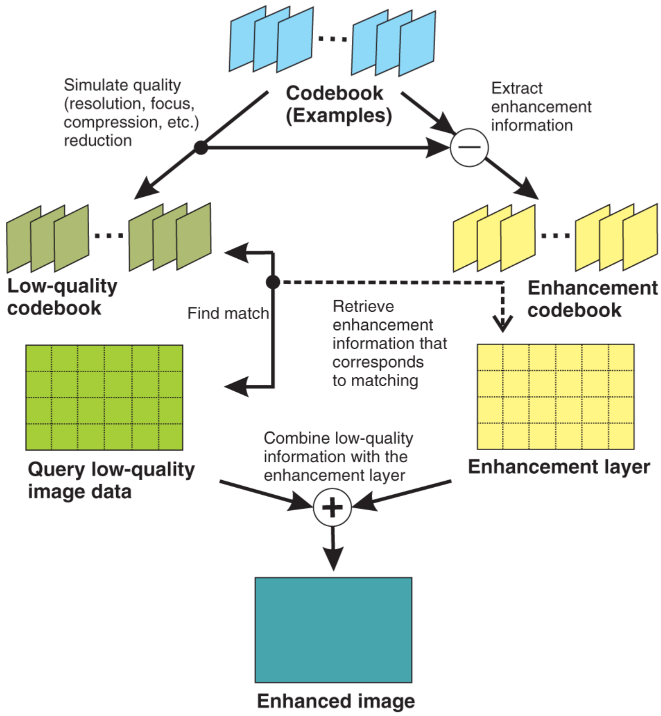

Fig. 2. General example-based enhancement scheme.

would be contained within a subspace of $\mathfrak{R}^N$, such that $\mathbf{y}$ can be viewed as a degraded (processed) version of $\mathbf{x}$.

Assume there is a codebook of example vectors $\{\mathbf{w}_i\}, \mathbf{w}_i \in \mathfrak{R}^N$. Each $\mathbf{w}_i$ is associated with its processed (degraded) version $F(\mathbf{w}_i)$. We then try to match the degraded versions, i.e. to match $\mathbf{y}$ to one of the $\{F(\mathbf{w}_i)\}$. The best match is then used to enhance $\mathbf{y}$. Let

$$\varepsilon_n = \|\mathbf{y} - F(\mathbf{w}_n)\|_p \tag{2}$$

be a matching-error L_p norm. If

$$\ell = \arg\min_n \varepsilon_n \tag{3}$$

then $\mathbf{x}$ is approximated by

$$\tilde{\mathbf{x}} = \mathbf{y} + \alpha(\varepsilon_\ell)(\mathbf{w}_\ell - F(\mathbf{w}_\ell)), \tag{4}$$

where α is a confidence weight that depends on ε_ℓ, i.e. on how good the match was. A graphical description of the above process is further depicted in Fig. 2 for redundancy and clarity.

Now assume we have a dictionary with a set of K codebooks $\mathbf{w}_{ki}$ $(1 \le k \le K)$ and we are to find a best match within each codebook. Then

$$\varepsilon_{kn} = \|\mathbf{y} - F(\mathbf{w}_{k,n})\|_p \tag{5}$$

$$\ell(k) = \arg\min_n \varepsilon_{k,n}, \tag{6}$$

and let the non-degraded version of the best-match block in the k-th codebook be $\mathbf{b}_k = \mathbf{w}_{k,\ell(k)}$. Hence, $\mathbf{x}$ is approximated by a weighted sum of the best matches within each codebook as

$$\tilde{\mathbf{x}} = \mathbf{y} + \frac{\sum_{k=1}^{K} \alpha(\varepsilon_{\ell(k)})(\mathbf{b}_k - F(\mathbf{b}_k))}{\sum_{k=1}^{K} \alpha(\varepsilon_{\ell(k)})}. \tag{7}$$

A set of weights that work well in many cases is $\alpha(\varepsilon_i) = 1/\varepsilon_i$, i.e. the smaller the distance of a matching block, the more important its contribution. Singularities in case of perfect matches can be easily circumvented by adding a very small quantity to the distance in (5).

The art involved in the process is to find codebooks that:

- are representative of the input data, in order to yield good results;
- can be easily re-populated to be useful in dynamic applications;
- allows speedy searches for the best match.

If a frame of a video sequence (or part thereof) is degraded, and if the neighboring frames (or frames temporally close) are not degraded, one might use them as references to create the codebooks as in Fig. 1. Let a block in the r-th frame be degraded, while K neighboring frames labeled f_1 through f_K are not. Furthermore, such frames are assumed to be very correlated with the r-th one. We populate each codebook $\{\mathbf{w}_{k,i}\}$ with patterns found in frame f_k.

The way that this can be easily reduced to practice is to carry a motion-estimation-like search for a block similar to the degraded target $\mathbf{y}$ within degraded versions of frames f_1 through f_K. We set $\mathbf{b}_k$ to be the non-degraded version of the best block match within frame f_k. Motion estimation is a pattern-matching search in temporally adjacent frames for matches with blocks that are spatially located in positions adjacent to the target block.

The algorithm for enhancing a degraded block $\mathbf{y}$ at the r-th frame, using frames f_1 through f_K as references is as following:

---

for i = 1 to K

  degrade frame f_i into $F(f_i)$

  carry motion estimation comparing $\mathbf{y}$ and $F(f_i)$

  retrieve best-matching block as $\mathbf{b}_i$

  set $\varepsilon_i = \|\mathbf{y} - F(\mathbf{b}_i)\|_p$

end for

Enhance $\mathbf{y}$ into $\mathbf{x}$ as

$$\mathbf{x} = \mathbf{y} + \frac{\sum_{i=1}^{K} \varepsilon_i^{-1}(\mathbf{b}_i - F(\mathbf{b}_i))}{\sum_{i=1}^{K} \varepsilon_i^{-1}}. \tag{8}$$

The method relies on the presence of non-degraded frames with similar content to the degraded areas. If the available non-degraded material is not very correlated it is advisable to extend the search over a large area so as to search over a diverse generic codebook.

III. Reproducible Degradation

The degradation of $\mathbf{x}$ into $\mathbf{y}$, modeled by F, can be part of a system design or an undesirable effect. In the first case, F is known. However, F may also be estimated from the actual degradation process. In any case, the process F needs to repeatable and reproduced over the codebooks (reference frames), as well. An example of a non-repeatable (and common) process is the addition of noise to the image, in which case one cannot

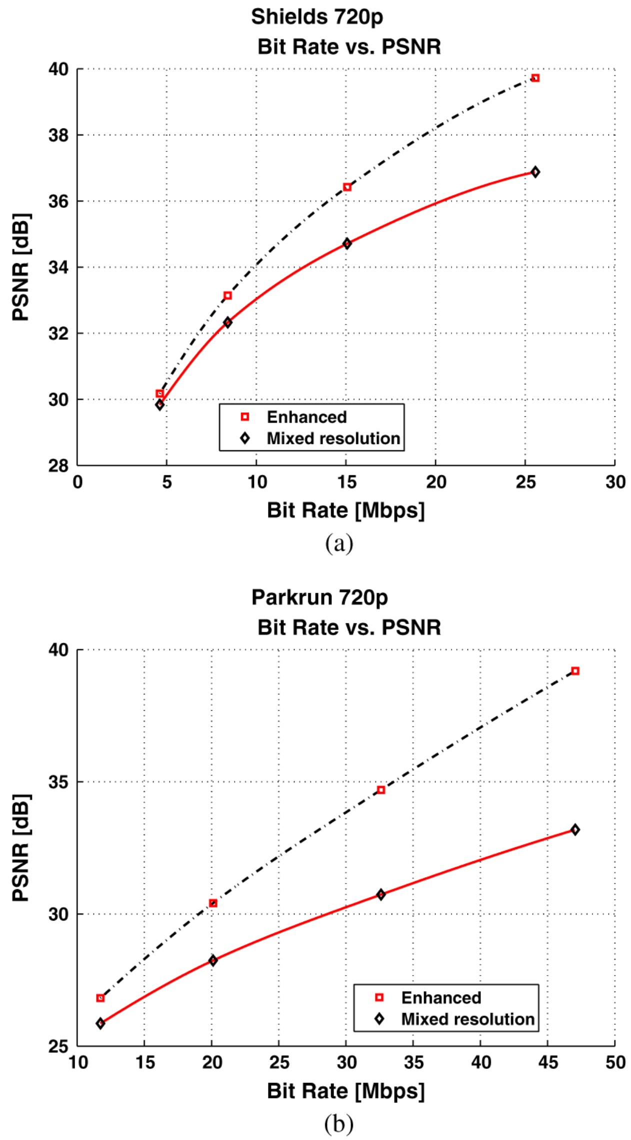

Fig. 3. Comparing resolution enhancement by example and interpolation for compressed mixed-resolution sequences Shields and Parkrun.

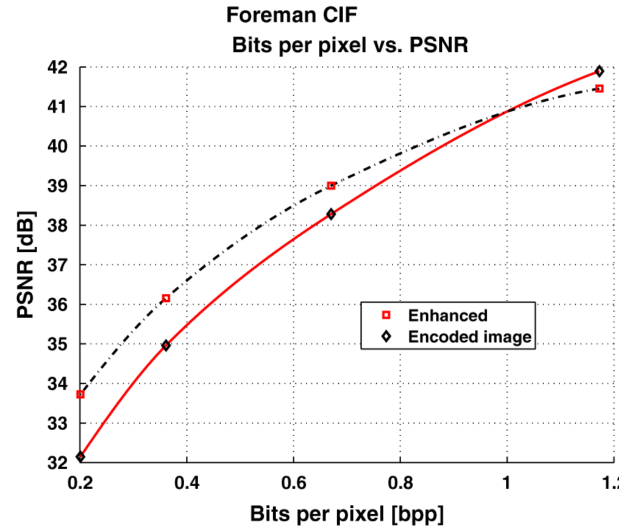

Fig. 4. Rate-distortion plots for evaluating the enhancement potential in case of degradation by lossy compression. 30-th frame in sequence Foreman (CIF) was compressed using H.264/AVC while uncompressed frames 29 and 31 were used as references.

add the same noise to the codebook entries in order to degrade them. Fortunately, there is also a possibility of enhancement by example in these cases.

Let $\mathbf{x}' = \mathbf{x} + \eta$ be the corrupted vector to be enhanced, where η is some noise vector, and let $F(\mathbf{x}')$ be the result of some denoising processing over $\mathbf{x}'$. Then,

$$F(\mathbf{x}') \approx F(\mathbf{x}) + G(\eta) \approx F(\mathbf{x}) \tag{9}$$

where G is some attenuation function over the noise. We assume some filtering operation for denoising would nearly remove the noise, despite degrading the image. As a result, we use $\mathbf{y} = F(\mathbf{x}') \approx F(\mathbf{x})$. In other words, in implementing (8) we assume:

- The degraded vector $\mathbf{y}$ is the denoised vector and not the noisy one.
- F is some repeatable efficient denoising operation.

IV. Application Scenarios and Experiments

In order to ascertain its capabilities, the framework will be applied to 4 forms of degradation: by (i) resolution reduction, (ii) quantization (compression), (iii) linear filtering and (iv) noise corruption. In the first case, F is a time-varying linear operator derived from the cascaded combination of decimation pre-filtering, down-sampling by a factor of M, up-sampling by a factor of M, and interpolation post-filtering. Hence, $\mathbf{y}$ is a blurred version of $\mathbf{x}$. In our approach the recovery of the low-resolution frames based on the high-resolution ones yields a process known as super-resolution by example [2]. As an experiment, every other frame of the video sequences was down-sampled by a factor of $M = 2$ in each direction. The sequence was compressed using H.264/AVC in Intra-only mode [14]–[16] before being decompressed and super-resolved (enhanced). Rate-distortion (RD) results are shown in Fig. 3 and compared to the trivial method of interpolating the low-resolution frames using Lanczos filters [17]. In Fig. 3, despite the good matching between frames patches, the quantization of high-frequency coefficients derived from the compression process reduces the information available for enhancement causing curves to be closer at lower bit-rates.

In the second case, F is the compression operation assuming a given encoder and a set of encoding parameters. Here, we used once more H.264/AVC in Intra mode. One interesting experiment is to compress one frame of a sequence and to enhance it using a temporally adjacent uncompressed frame as references for the codebook, whose RD curves are shown in Fig. 4. In another experiment, even numbered frames of a sequence are compressed with AVC's quantization parameter Q while the odd ones are compressed with $Q+6$ (lower quality). RD results are shown in Fig. 5.

In the third case, the degradation is caused by filtering, and in our tests, the operator F is the convolution with a square Gaussian kernel. In our experiments, 300-frame sequences are blurred (defocused) where a non-blurred (focused) frame periodically occurs at every 30 frames. Table I shows the objective results (average PSNR) of the proposed enhancement applied to a few video sequences, which indicates an average distortion reduction of 8.79 dB.

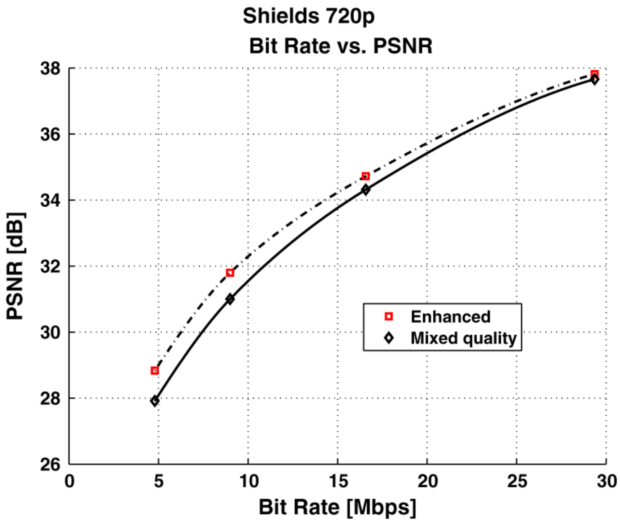

Fig. 5. RD results for the enhancement of mixed-quality sequence Shields.

TABLE I
AVERAGE PSNRS (IN DB) AND SSIMS OF DEGRADED FRAMES: UNFOCUSED WITH AN 8×8 GAUSSIAN FILTER BEFORE AND AFTER THE ENHANCEMENT PROCESS

Sequence	Blurred	Enhanced
Foreman	25.03 (0.716)	31.30 (0.871)
News	21.81 (0.705)	33.17 (0.953)
Mobile	17.73 (0.336)	24.61 (0.876)
Hall	22.07 (0.710)	33.18 (0.938)
Container	21.29 (0.606)	33.92 (0.903)
Shields	25.62 (0.648)	30.24 (0.803)
Mobcal	22.62 (0.517)	31.28 (0.898)
Average	22.31 (0.605)	31.10 (0.892)

In the fourth and last case, degradation comes from corruption by noise, which in our experiments was set to "salt and pepper" noise added to 2% of the pixels. As discussed, the operator F is the median filter, which is often used to deal with this kind of noise. In our experiments, example frames are noiseless and periodically occur at every 30 frames of the 300-frame sequences, and we used 5×5 median filters. For subjective comparisons, Fig. 6 depicts a degraded (noisy) frame of sequence "Shields" after processing with a median filtered and with the proposed enhancement using examples. Table II gives test results applied to other video sequences. Results indicate an average 9.81 dB improvement over the degraded images, and a 6.45 dB average distortion improvement over the degraded frames processed with median-filters.

Fig. 6. Region of the 16th frame of the video sequence *Shields* corrupted by '*salt and pepper*' noise: top, filtered (with a median filter); bottom, enhanced using median-filtered examples.

TABLE II
AVERAGE PSNR (IN DB) AND SSIM OF DEGRADED AND ENHANCED FRAMES AFTER CORRUPTION BY SALT-AND-PEPPER NOISE

Sequence	Noisy	Median-filtered	Enhanced
Foreman	22.14 (0.620)	29.48 (0.841)	33.44 (0.913)
News	21.86 (0.621)	26.01 (0.881)	35.25 (0.973)
Mobile	22.05 (0.782)	19.43 (0.553)	24.33 (0.867)
Hall	22.37 (0.605)	25.80 (0.847)	34.64 (0.950)
Container	22.41 (0.618)	23.76 (0.722)	33.51 (0.905)
Shields	21.78 (0.644)	29.01 (0.808)	31.98 (0.866)
Mobcal	22.25 (0.675)	24.85 (0.681)	30.36 (0.889)
Average	22.12 (0.652)	25.48 (0.762)	31.93 (0.909)

V. CONCLUSIONS

In this paper, we discuss an example-based approach for the general enhancement of degraded video frames wherein there are non-degraded parts of the video from where to build the dictionary. The dictionary-building and search processes reduce to motion estimation among the degraded and the reference frames. The image degradation has to originate from a "repeatable" process, and in the case of non-repeatable noisy operations there should be a repeatable denoising process from where to build the "degraded" reference. The method is a generalization of the example-based super-resolution approach for mixed-resolution video. Its complexity is estimated to be similar to that of motion estimation algorithms.

Results are consistent and point to significant gains in face of many forms of degradations. The results highlight the potential applicability of the method to many situations.

Further studies include proposing the use of a non-reference-image-quality estimator in order to control the amount of detail information to be added to the degraded frame. We also plan to study possible flickering effects.

References

[1] W. Freeman, T. Jones, and E. Pasztor, "Example-based super-resolution," *IEEE Comput. Graph. Applicat.*, vol. 22, no. 2, pp. 56–65, Apr. 2002.

[2] E. M. Hung, R. L. de Queiroz, F. Brandi, K. F. Oliveira, and D. Mukherjee, "Video super-resolution using codebooks derived from key-frames," *IEEE Trans. Circuits Systems Video Technol.*, vol. 22, no. 9, pp. 1321–1331, Sep. 2012.

[3] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 21–36, May 2003.

[4] G. Cristobal, E. Gil, F. Sroubek, J. Flusser, C. Miravet, and F. B. Rodriguez, "Superresolution imaging: A survey of current techniques," in *Proc. SPIE*, 2008, vol. 7074, pp. 70 740C–1–70 740C–18.

[5] A. K. Katsaggelos, R. Molina, and J. Mateos, *Super Resolution of Images and Video (Synthesis Lectures on Image, Video, and Multimedia Processing)*. San Rafael, CA, USA: Morgan and Claypool, 2006.

[6] S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, pp. 1013–1027, Jun. 1990.

[7] S. P. Kim and W. Y. Su, "Recursive high-resolution reconstruction of blurried multiframe images," *IEEE Trans. Image Process.*, vol. 2, pp. 534–539, Oct. 1993.

[8] B. C. Song, S. C. Jeong, and Y. Choi, "Video super-resolution algorithm using bi-directional overlapped block motion compensation and on-the-fly dictionary training," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 3, pp. 274–285, Mar. 2011.

[9] D. T. Vo and T. Q. Nguyen, "Quality enhancement for motion JPEG using temporal redundancies," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, pp. 609–619, May 2008.

[10] D. T. Vo and T. Q. Nguyen, "Optimal motion compensated spatio-temporal filter for quality enhancement of H.264/AVC compressed sequences," in *Proc. IEEE Int. Conf. Image Processing*, Nov. 2009, pp. 3173–3176.

[11] S. C. Jeong, T. H. Lee, B. C. Song, Y. Lee, and Y. Choi, "Video deblurring algorithm using an adjacent unblurred frame," *Vis. Commun. Image Process.*, Nov. 2011.

[12] S. Kanumuri, O. G. Guleryuz, M. R. Civanlar, A. Fujibayashi, and C. S. Boon, "Temporal flicker reduction and denoising in video using sparse directional transforms," in *Proc. SPIE Conf. Applications of Digital Image Processing XXXI*, 2008.

[13] J. Y. H. Yue, X. Sun, and F. Wu, "Cloud-based image coding for mobile devices–toward thousands to one compression," *IEEE Trans. Multimedia*, vol. 15, no. 4, pp. 845–857, Jun. 2013.

[14] T. Wiegand, G. Sullivan, G. Bjoontegaard, and A. Luthra, "Overview of the h.264 video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 560–576, Jul. 2003.

[15] M. Yang, M. Trifas, G. Xiong, and J. Rogers, "H.264/AVC intra-only coding (iAVC) techniques for video over wireless networks," in *SPIE Int. Conf. Electronic Imaging*, Jan. 2009, vol. 7256, pp. 725 608–1–725 608–10.

[16] T. Wedi, H. Ohtaka, J. Wus, and S. Sekiguchi, Intra-only H.264/AVC profiles for professional applications Joint Video Team (JVT), Hangzhou, China, Tech. Rep. JVT-U120, Oct. 2006.

[17] C. E. Duchon, "Lanczos filtering in one and two dimensions," *J. Appl. Meteorol.*, vol. 18, no. 8, pp. 1016–1022, Aug. 1979.