MEASURING NOISE CORRELATION FOR IMPROVED VIDEO DENOISING

Anil Kokaram, Damien Kelly, Hugh Denman, Andrew Crawford*

Chrome Media Group, Google Inc, 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA

ABSTRACT

The vast majority of previous work in noise reduction for visual media has assumed uncorrelated, white, noise sources. In practice this is almost always violated by real media. Film grain noise is never white, and this paper highlights that the same applies to almost all consumer video content. We therefore present an algorithm for measuring the spatial and temporal spectral density of noise in archived video content, be it consumer digital camera or film orginated. As an example of how this information can be used for video denoising, the spectral density is then used for spatio-temporal noise reduction in the Fourier frequency domain. Results show improved performance for noise reduction in an easily pipelined system.

***Index Terms*—** Noise Measurement, Video Denoising, Denoising, Noise Reduction, Video Noise Reduction, Wiener Filter, 3D Fourier Transform

1. INTRODUCTION

Noise reduction for video sequences is a well studied signal processing topic. The core problem is always to remove noise without affecting image details. Early practical video denoisers were developed by Dubois, Dennis et al [1] in the 1980's, based on motion compensated temporal filtering with no spatial information. Provided the motion compensation was accurate, this held the potential for excellent preservation of details. However the dirty window side effect naturally led to the design of spatio-temporal noise reduction filters introduced by Katsagellos, Sezan, Lagendijk et al [2] by the mid 1990's. Transform domain noise reduction for video was introduced toward the end of that decade involving transform coefficient shrinkage of some kind in the fourier frequency and wavelet domains [3, 4]. This held the potential for much stronger noise reduction power with better detail preservation and indeed these techniques have been used in commercial denoisers[1] for some time. The seminal work of Efros et al at the turn of the century enabled researchers to build image models from the redundancy in the image itself and this led to the now popular non-local means algorithm for noise reduction. It is only relatively recently however that video denoising schemes based on this idea have arisen [5, 6]. The combination of sparse coding ideas and non-local means has now led to the general notion of shrinkage in an adaptive basis space tuned to images and video e.g. Sapiro et al [7]. The key improvement introduced by these algorithms has been the potential for exact preservation of strong image details together with very strong noise reduction.

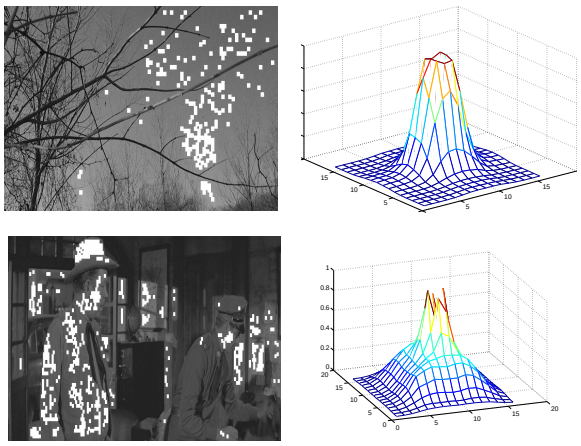

Fig. 1. *Two examples of typical footage (Top: Consumer footage, Bottom: 2K Film scan), showing on the left the blocks over which the N-PSD (shown on the right) was measured. The maximum noise PSD aplitude is scaled to 1.*

There is therefore no shortage of ideas for denoising consumer video footage. However, none of the previous work has addressed one important issue: noise in real video is hardly ever uncorrelated. While it is well known that film grain noise is not white, the work of Buades et al [6] attempts to show that noise in digital cameras is signal dependent, additive and white. However this ignores the reality of the entire processing chain of consumer capture and upload which must introduce spatial correlation. Immediately after the sensor for instance, there is a de-mosaicing step that upsamples the colour channels before compression. Particularly in the case of online video repositories, the numerous transcoding and upsampling/downsampling stages that are applied at *every* stage of upload and download imply that any noise is now spatially correlated and combined with compression artefacts. In this paper we lump all of these defects into our definition of noise as we attempt to address real picture degradation.

*anilkokaram@google.com

[1]Neat Video, The Foundry, RedGiantSoftware, Snell and Wilcox

Figure 1 shows the power spectral density (psd) of noise measured in a 720p consumer clip and a 2K professional film scan. In both cases the psd is not flat and hence the noise is *not* white. We have analysed over 100,000 clips (each over 10 sec in length) from various consumer sources and the noise is *not* white. By ignoring this correlation structure, current denoising algorithms run the risk of attenuating more of the signal spectrum (regardless of basis) than is necessary. This leads to more blurring of details than is desirable in a practical system. Addtionally, non-local means and sparse denoising concentrates on modelling the underlying picture in some implicit sense, but by capturing the noise structure we are also able to incorporate information about the noise. We therefore turn a one class problem into a much more optimistic two class problem. Furthermore, in sparse denoising, noise correlation causes a residual noise component in the adaptive bases computed in the sparse representation.

We propose in this paper to measure the noise correlation structure from the degraded video itself. The idea is to use a low complexity step to select a number of candidate image patches or blocks which contain noise-like texture and then process these candidates in a more complex second step. This second step allows the correlation structure to be expressed in the right form for the denoiser that is to be used. A wavelet denoiser would require measurement of the noise variance in each subband, a sparse shrinkage denoiser would require the noise power in each of the basis functions. As an example, in this paper we use the overlapped block 3D Fourier Frequency denoiser first introduced in [3]. The denoiser is easy to understand and has many similarities to the emerging ideas in sparse shrinkage denoising, without the computational complexity. In fact it is the precursor to the block matching based denoiser of Rusanov et al [8]. We show that by simply measuring the noise structure successfully, we can produce state of the art results. We next state our assumptions about noise and the noise degradation model, then go on to present the noise measurement subsystem.

Noise Model: Given the clean pixel intensity $I_n(h,k)$, at site (h,k) in frame n, we assume additive noise such that $G_n(h,k) = I_n(h,k) + \nu_n(h,k)$ where the observed dirty pixel intensity is $G_n(h,k)$ and the additive noise is $\nu_n(h,k)$. That noise is assumed to be correlated and for the purposes of this paper its correlation structure is expressed through its 3D Power Spectral Density in the 3D Fourier Frequency domain. See Appendix A for further definitions. This is the appropriate form for denoising in the Fourier Domain with the overlapped block Wiener Filter [3]. However the correlation structure can be expressed in any other form for use with other denoisers. The amplitude of the noise PSD (N-PSD) in the qth plane of the (r,s)th frequency bin is denoted by $P_q^N(r,s)$. We further assume that the N-PSD is spatially and temporally symmetric i.e. there is no directional or coherent textural structure. In this paper we do not allow for signal dependent noise. Modifications that allow for signal dependent noise will be presented in future work.

2. NOISE SPECTRAL DENSITY MEASUREMENT

We measure noise by first selecting patches in the current frame which are likely to contain noise added to low complexity image texture. We then use the residual arising from a low-order image model of those patches to estimate the noise statistics from a robustly chosen subset of patches. Finally, from the corresponding motion compensated volumes we generate an estimate of the noise power spectral density. Patches are $B \times B$ pixels, and are visited with an overlap of 50% in the image. Given a patch $G_n(h,k)$ in frame n, we define the corresponding motion compensated patches in the previous and next frames as $G'_{n-1}(h,k)$, $G'_{n+1}(h,k)$ respectively. We compensate with translational global motion which reduces the computational load of the process without affecting quality.

Selecting candidate noise patches: Suitable candidates contain no dominant edge or directional gradient structure. Our candidate patch selector is therefore based on a modification of the ideas behind corner detection. However texture analysis is confused by noise in the image, which leads to the paradox that we need to know the noise to estimate the appropriate patches for noise measurement. To resolve this, we low pass filter the image under consideration with a light gaussian filter (variance of 2 here) to reduce the influence of noise to some extent. Statistics gathered over patches then gives further robustness.

Given the filtered patch $G_n^f(h,k)$ and associated gradients in the horizontal and vertical directions represented as column vectors, $[\mathbf{g}_x,\ \mathbf{g}_y]$, the well known gradient covariance matrix is $\mathbf{C}_{gg} = [\mathbf{g}'_x\mathbf{g}_x \quad \mathbf{g}'_x\mathbf{g}_y;\quad \mathbf{g}'_x\mathbf{g}_y \quad \mathbf{g}'_y\mathbf{g}_y]$. Defining the maximum and minimum eigenvalues of this matrix as $\lambda_1,\ \lambda_2$, appropriate candidate patches are selected when the following conditions are met: $|\mathbf{C}_{gg}| < T_d;\ \ Tr(\mathbf{C}_{gg}) < T_g$; and $\lambda_1/\lambda_2 < T_l$. We use T_l to control the amount of dominant directional information in a patch, T_g controls the total amount of gradient energy and T_d is a proxy for the cornerness measure. Our experiments use $T_d = 0.04,\ T_g = 0.9,\ T_l = 30$. Figure 1 shows typical patches selected with this process in one example frame ($B = 16$).

Finally, for patches to yield a robust spectral measure, the corresponding patches in the surrounding frames must not contain texture information. Hence we reject any patches which have large motion compensated DFDs (> 10 in our experiments below); where DFD is $(1/B^2)\sum_{h,k}|G_n(h,k) - G'_{n+1}(h,k)|$, and similarly for the previous frame.

Patch Analysis for Noise: A planar model with coefficients a_{0-3}, $\hat{G}_n(h,k) = a_0 + a_1h + a_2k + a_3hk$, for the underlying image data is fit to each candidate patch G_n^f (using LS) in the current frame. Since patches are selected to have little underlying image texture information, the residual error $e_n(h,k) = G_n(h,k) - \hat{G}_n(h,k)$ provides a possible estimate

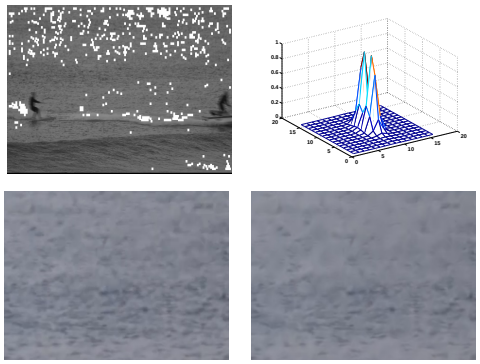

Fig. 2. *A failure case. The original clip in the top left yields the measured N-PSD shown on the right using the blocks as indicated. $c_s = 7.3$ clearly indicates asymmetry in the N-PSD. A zoom on the original frame (bottom left) shows some of the wave texture removed in the denoised output (right).*

of the noise variance σ^2_{ee} in each patch. Note we use the actual degradaded patch G_n for the residual. Clearly not all the patches will be well modelled by a plane, and those will most likely contain some image texture information. To robustly select those patches that are most likely to represent noise, we generate a histogram of σ^2_{ee} and those patches that lie within the lower quartile are assumed to contain only noise.

Each of the residual patches $e_n(h,k)$ are then combined with their counterparts in frames $n-1,\ n+1$ and used to generate a set of final candidate N-PSD's, $\mathcal{N}_p$ as in Appendix A. Each example in $\mathcal{N}_p$ arises from the correlation structure of the underlying stochastic noise generation process ν. The MMSE estimate for the N-PSD is the average of these PSDs. In practice we find that the median of each frequency bin across the set $\mathcal{N}_p$ gives a more useful N-PSD $\hat{P^N_q}(r,s)$ for denoising.

Validating the N-PSD: The noise measurement algorithm as described above can still confuse legitimate image texture with undesired noise e.g. carpet and waves. Figure 2 shows an example. Recall that our definition requires noise to be non-directional both in space and time. Hence the N-PSD is circularly symmetric and the noise power should not vary with time. We therefore define two confidence measures, $c_s,\ c_t$, that assess the shape of the N-PSD. Spatial symmetry is captured with $c_s(n)$, by measuring the ratio of energies in a narrow band of horizontal and vertical frequencies in $P_n(h,k)$, as follows.

$$\Delta(q) = \sum_{r=0}^{1}\sum_{s=0}^{1} P_q(B/2+r, B/2+s)$$

$$c_s(q) = \frac{\sum_{r=0}^{B-1}\sum_{s=0}^{1} P_n(r, B/2+s) - \Delta(q)}{\sum_{r=0}^{1}\sum_{s=0}^{B-1} P_q(B/2+r, s) - \Delta(q)}$$

where $\Delta(q)$ measures the energy around the lowest frequencies of the spectral plane P_q. Removing $\Delta(q)$ is important. That energy tends to compose the bulk of the energy of the PSD in that plane, hence would dominate the ratio and desensitise the metric. Valid Noise PSDs show $c_s \approx 1$, and we use $c_s < T_s$ where $T_s = 1.25$.

Temporal symmetry is measured with c_t as follows.

$$c_t = \frac{\sum_{r=0}^{B-1}\sum_{s=0}^{B-1} P_0(r,s)}{\sum_{r=0}^{B-1}\sum_{s=0}^{B-1} P_1(r,s)}$$

If the noise PSD is valid the power of the noise PSD in each spectral plane should be approximately the same, hence $c_t \max(c_t, 1/c_t) < T_t$. In our experiments $T_t = 3.0$.

3. THE NOISE REDUCTION SYSTEM

The overlapped block 3D Wiener Filter [3] is used for denoising. Given our estimated N-PSD $P^N_q(r,s)$, the filter $H_q(r,s)$ in the spectral domain can be modified from [3] as follows.

$$H_q(r,s) = \frac{f(P^G_q(r,s) - P^N_q(r,s))}{P^G_q(r,s)}$$

$$\text{where } f(x) = \begin{cases} P^G_q(r,s) - P^N_q(r,s) & P^G_q(r,s) > \beta P^N_q(r,s) \\ \frac{\beta-1}{\beta} P^N_q(r,s)) & Otherwise \end{cases}$$

Here $\beta = 1.1$ prevents discontinuities in the spectrum of the output signal after filtering. See [3] for further details.

For fully automated processing of entire movies, the system has to deal with shot transitions. In each shot a separate measurement of the N-PSD must be made since the noise can change between shots. We use a histogram based shot cut detector to make another noise measurement. If less than K candidate patches are selected in a frame for noise measurement, the process is repeated in the next frame. We only update N-PSD between shots if the change in noise energy is significant (25% here).

4. RESULTS AND FINAL COMMENTS

Figure 3 shows frames from two example sequences, one is a typical consumer video upload (Top) and the other from a raw 2K scan of an archived film (hence no compression artefacts). The images show that using $\hat{P^N_q}(r,s)$, there is better detail preservation in the denoised output. We can use the metric Q introduced by Milanfar et al [9] to give some quantitative indication about the tradeoff between detail loss and noise. For the top row $Q = 46.3,\ 37.2$ when using the N-PSD and AWGN respectively. For the bottom the metric is $12.7,\ 10.3$ respectively. Thus in both cases Q reinforces the visual comparison. Also in both cases, $c_t \approx c_s \approx 1$. Figure 2 shows a failure case that is detected by the symmetry measures presented earlier $c_t = 1.5, c_s = 7.3$. In our pipelined system, this kind of material is left untouched.

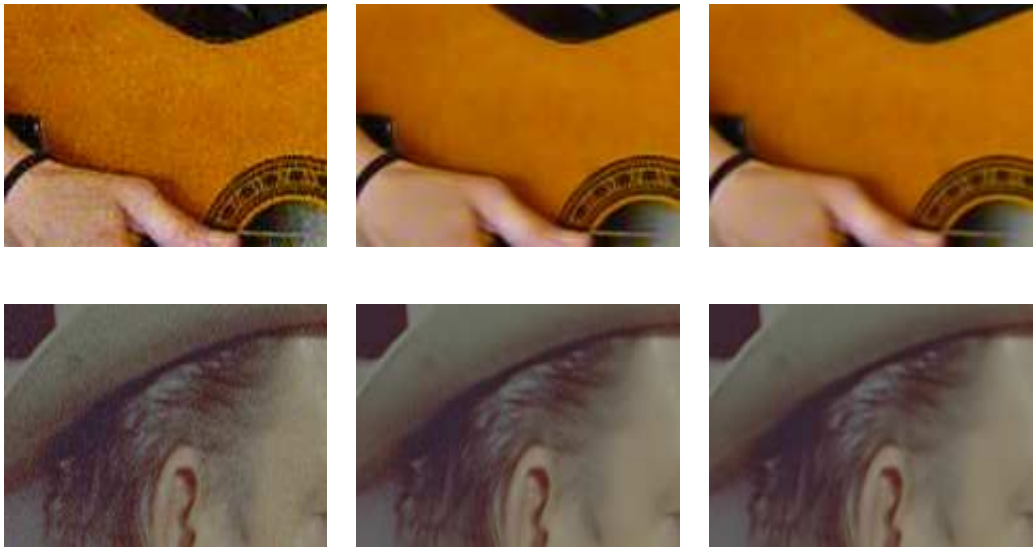

Fig. 3. *Two zooms on examples of typical footage. Top row: Consumer footage showing heavy noise, Bottom Row: A 2K film scan. Left to right shows Original (degraded) footage, denoising using the measured N-PSD, denoising assuming AWGN of the same variance. In all cases using the measured N-PSD gives much better detail preservation with the same noise reduction power in flat areas.*

We have presented a system for measuring the correlation inherent in noise in real media, together with a mechanism for detecting failures. Our examples show that it is relatively simple to incorporate this information into spectral domain denoisers, and does yield improved detail preservation. Our project page will be updated with our complete set of examples in due course. It is however not clear how non-local means denoising can be informed with correlation information and we consider this in future work.

A. THE 3D FFT

The 3D FFT is defined as follows.

$$F_q(r,s) = \sum_{n=0}^{N-1} \omega_3^{nq} \sum_{h=0}^{B-1} \omega_2^{rh} \sum_{k=0}^{B-1} G_t(h,k)\omega_1^{sk}$$

where $G_t(h,k)$ is the intensity at site h,k in frame t, and $F_q(r,s)$ is the 3D spectral coefficient in plane q at frequency bin (r,s). The 3rd FT is taken over N frames of G_t, and the patch in G_t is $B \times B$ pixels. We assume that spectral coefficients $F_q(r,s)$ in each plane q are rearranged so that the DC coefficient is at $(B/2+1, B/2+1)$.(`fftshift` in Matlab). The power spectral density $P_q(r,s)$ is then $|F_q(r,s)|^2$.

B. REFERENCES

[1] E. Dubois and S. Sabri, "Noise reduction in image sequences using motion compensated temporal filtering.," *IEEE Transactions on Communications*, vol. 32, pp. 826–831, July 1984.

[2] J. Brailean, R. Kliehorst, S. Efstratiadis, A. Katsaggelos, and R. Lagendijk, "Noise reduction filters for dynamic lmage sequences: A review," *IEEE Proceedings*, vol. 83, no. 9, Sept 1995.

[3] A. Kokaram, "3D wiener filtering for noise suppression in motion picture sequences using overlapped processing.," in *Signal Processing V, Theories and Applications*, September 1994, pp. 1780–1783.

[4] Peter M. B. Van Roosmalen, S. J. P. Westen, R. L. Lagendijk, and J. Biemond, "Noise-reduction for image sequences using an oriented pyramid thresholding technique," in *IEEE International Conference on Image Processing*. 1996, vol. 1, pp. 375–378, IEEE.

[5] Mona Mahmoudi and Guillermo Sapiro, "Fast image and video denoising via nonlocal means of similar neighbourhoods," *IEEE Signal Processing Letters*, vol. 12, no. 12, December 2005.

[6] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel, "Nonlocal image and movie denoising," *International Journal of Computer Vision*, vol. 76, pp. 123–139, 2008.

[7] Matan Prottera and Michael Elad, "Sparse and redundant representations and motion-estimation-free algorithm for video denoising," in *Proc. SPIE Electronic Imaging*, 2007, vol. 6701.

[8] D. Rusanovskyy and K. Egioazarian, "Video denoising algorithm in sliding 3d dct domain," in *Advanced Concepts for Intelligent Video Systems*, Sept 2005, vol. 6064A-30.

[9] Xiang Zhu and Peymann Milanfar, "A no-reference image content metric and its application to denoising," Sept 2010, pp. 1145–1148.