Image de-Noising by Selective Filtering Based on Double-Shot Pictures

Dinei A. Florêncio

Microsoft Research, One Microsoft Way, Redmond, WA 98052

***Abstract*—The quality of photographs is often reduced by sensor noise. This was a problem with film cameras, and is still a problem with current CCD and CMOS sensors, particularly in low-lighting conditions. De-noising techniques do not always perform satisfactorily. Typical de-noise techniques reduce the sharpness of the image. In this paper we propose a new de-noising technique, which is based on a dual-shot technique. The proposed algorithm is based on selectively removing high frequencies that do not correlate well between the frames. The algorithm borrows from video processing noise-removal techniques, but the final picture is derived from filtering a single shot, avoiding double-contouring and other artifacts that may happen with video techniques. While the decision is made based on both frames, the filtering itself is done using exclusively one of the frames. For this reason, the second (auxiliary) shot may be of much lower quality.**

Keywords - CCD; noise reduction; sensor noise; low-light; de-noising; double-shot; photography.

I. Introduction

Most photographs are contaminated by noise. This was a problem with film cameras, and still is a problem with digital cameras. In fact, noise is essentially intrinsic to image capture devices, including CCD and CMOS sensors. While this noise may be negligible for high-contrast, well lit scenes, it may become significant for dark or low contrast scenes. Several techniques have been proposed for reducing sensor noise. A common approach is to low-pass the image, since typical CCD noise is uncorrelated from pixel to pixel. To limit the smoothing of the high frequency content of the image, a special filter is often used, preserving the high frequency at the edges, whenever one is detected. Examples of such techniques include directional filtering, and bi-lateral filter. Other techniques try to preserve the edges by using median or other non-linear techniques.

A more recent approach tries to obtain information from multiple shots of the same scene, in order to improve the quality of the noise removal. In many respects, these techniques are similar to noise-removal or pre-processing techniques used in video, where pre-processing involves a temporal-filter, based on motion-compensated frames [1-3]. Nevertheless, areas that are not perfectly motion compensated will introduce artifacts (or "phantoms). This type of artifact may be fine for video, where a frame will be displayed for only a fraction of a second. Nevertheless it would be completely unacceptable for still images. In this paper, we propose a new filtering method, based on double shot of the same scene. With digital cameras, the actual cost of the double shot is low, and the double shot approach has been proposed before for different applications [5]. The proposed algorithm reduces the camera noise without removing legitimate texture, smoothing edges, or introducing double-contour artifacts.

II. Can We Differentiate Between Sensor Noise and Actual Signal?

How exactly can we remove the camera noise without affecting too much of the picture? We first note that camera noise, while spreading over the whole frequency range, is most important at higher frequencies, where it is expensive to encode, and often more intense than the high frequency contents of the desired signal itself. A simplistic solution would be therefore to low-pass the image. Nevertheless, doing so would also significantly reduce the sharpness of the image, ending up with a lower quality signal. To solve this problem, we propose using an auxiliary picture, which can be used as reference to check whether a certain detail comes from the actual scene or if it is mostly camera noise. By using an auxiliary picture of the same scene, these two sources of high-frequency can be differentiated, *even if they have the same spectrum*. More specifically, the high frequency content of the image is going to be consistent across both shots, while the camera noise will be independent. If needed, small changes or camera motion can be handled by motion compensation.

III. Spatial Vs.Temporal Filtering

By having access to multiple frames, the problem is reduced to almost the same problem as that of pre-processing video. As we mentioned in the previous section, most of the information to be removed consist of high frequencies. Therefore, one could try to attenuate the undesired high frequencies by simple spatial low-pass filtering. Nevertheless, only temporal filtering will attenuate camera noise while at same time preserving the high frequency contents of the actual image. For this reason, most pre-processing methods generally involve some sort of temporal filtering. The key disadvantage of these temporal filtering methods is the phantoms (double-contours) generated in the image, even when motion compensated filtering is used. Furthermore, to obtain enough noise attenuation, a large number of frames would be needed, which may not be practical.

In [7], we propose a video pre-processing algorithm where – in essence – the decision about filtering or not comes from the temporal analysis, but the filtering itself is done in the spatial domain. We now extend that algorithm to the case of still images, based on a double-shot approach.

IV. THE PROPOSED ALGORITHM

The proposed method is based on the idea that sensor noise is not correlated between successive shots, and therefore cannot be predicted from one shot to the next. Nevertheless, different from the video case, where every frame is going to encoded, in the photograph case, only one of the shots is of interest. For this reason, is very likely that one of the shots may be a low-quality one. For example, it may have a lower-power flash, or a lower exposure time. We therefore also re-match the picture for lighting and contrast compensation. After that, we compare the high frequency content of the main frame with the high frequency content of the motion compensation residual. High frequencies that are present in the image, but not in the motion compensated residual will represent content derived from the scene, and should therefore be preserved. All other high frequency should be attenuated, since it will correspond mostly to camera noise. Figure 1 presents a high level diagram of the proposed method. We first compensate the auxiliary picture to produce a frame with similar lighting to the main picture. Then, we produce a motion compensated (MC) frame, and a high pass version of both the original frame and the MC residual. High frequency "energy" content is compared between these two, and a ratio based on these two energy values determines how much to attenuate the high frequency contents of the original frame.

More precisely, given a Picture P, we produce high pass version of the frame, Php, given by Php = P*hh, where hh is a high pass FIR filter. In our simulations we used a separable 5-tap filter with a smooth transition around .3π, but other filter could be used, tuned to the amount of smoothness desired, or to computational constrains. This high pass frame is then squared and low pass filtered (we used a 7×7 tap separable linear filter). This produces an estimate of the amount of high frequency in each region of original frame, which we call EF.

A motion compensated picture residual R is also produced. Depending on computational and other requirements, this motion compensation can be either a camera-motion only (i.e., one single motion vector for the whole image), or a full-fledged motion compensation, which would also compensate for subject motion. An estimate of the high frequency content in the MC residual ER is then produced by following the same steps used to produce EF. We now have an estimate of the high frequency content of the original picture (EF), and of the (motion compensated) frame difference.

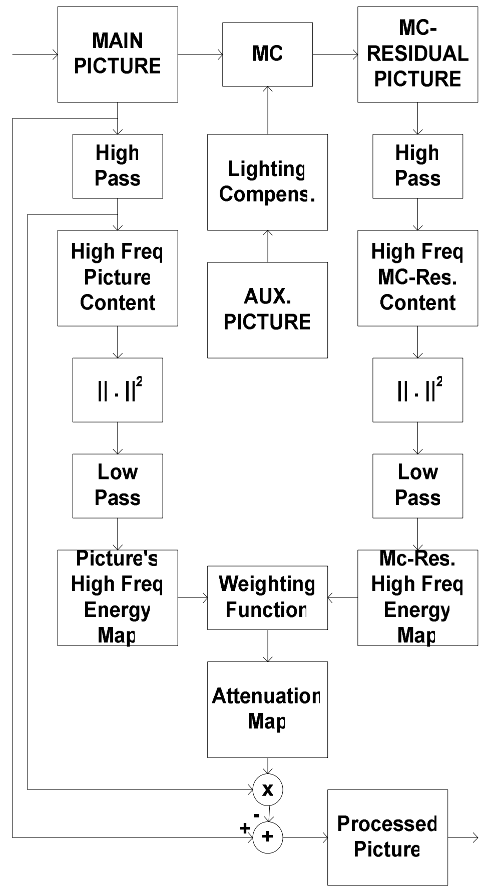

Figure 1 – *overall diagram of the proposed algorithm.*

We then obtain an attenuation map by computing, for each pixel:

$$ATT = \max(0, \min(1, G1.(\mathrm{ER}/(\mathrm{ER}+\mathrm{EF}) - \mathrm{B1}))) \quad (1)$$

where G1 is a gain factor and B1 a bias factor. To understand the theoretical values of G1 and B1, let us analyze two extreme

Figure 2 – *Main Picture*

Figure 3 – *Processed Picture*

Figure 4 - *High Frequency content of original picture*

Figure 5 - *High frequency content of processed image*

cases. First, suppose that all high frequencies are from the image. In this case, the MC residual will be zero (as long as the motion compensation works). Since all high frequency content is actually from the image, we do not want to attenuate any of it, and therefore the theoretical value for B1 is zero. A higher value may be used to preserve more of the high frequencies, since the MC tends to never exactly match, due to the motion vector precision, or other factors. Now, let's examine the other extreme. If the desired image is completely flat, all high frequencies on both the original frame and on the MC residual will be due to camera noise. If we assume that the MC did not track this camera noise, the energy level in the MC should be around twice the energy level in the original frame. Therefore, setting G1=1.5 will yield ATT=1, and therefore remove all high frequencies from this region of the image, as we would expect. A higher value of G1 will increase noise attenuation, and may also be used to compensate the fact that the MC may track some of the camera noise. Or, a lower value may be used to be conservative and preserve more of the high frequencies. In our experiments, we have set to G1 = 1.5, and B1 = 0.

The final frame is the obtained by subtracting the (attenuated) high frequency content Php from the original frame P, i.e.:

$$OP = P - ATT \cdot Php \qquad (2)$$

where OP is the output picture.

Figure 6 - *Attenuation map (whiter = attenuate more)*

Note that in the above explanation (and in Figures 3 to 7) we use only one high frequency band. This is not a constraint of the algorithm; in fact, we could use as many bands as we desire. In particular, one could use a wavelet or other decomposition, and produce an attenuation map for each subband.

IV. RESULTS

Figures 2-8 illustrate some of the results obtained by applying the algorithm. We run tests on two types of data. First, since these are easily available, we tested our algorithm on a video sequence. We then tested with data acquired by a digital still camera under low-lighting conditions. Let us first look at the video images. Figure 2 is the original frame 50 of the MPEG-4 test sequence SEAN, while Figure 4 represents the high frequency content in that same frame. Figure 6 shows the attenuation map, which highlights regions where the high frequency content was not tracked by the motion compensation, and should therefore be attenuated. Note that all textured regions of the background (e.g. the plants and the textured columns) appear dark in the attenuation map, meaning they will have their detail information preserved. The same is true of regions that are well tracked by the MC, like the dark suit contour, or the tie. In contrast, flat regions of the background (where high frequencies are dominated by camera noise) will be low-passed, like the sofa or the wall. This attenuation of the undesired high frequencies can be observed by comparing Figures 4 and 5, which show the high frequency content of the original and output (processed) images, respectively. Figures 3 show the final image. Figure 7 show a side-by-side comparison between the original and processed images, zoomed on the region around the sofa. Note the preservation of the texture and sharpness in textured regions (the column, the plants), but the absence of sensor noise in non-textured regions (e.g., the suit, the wall, and the sofa).

Finally, Figure 8 shows a more important case: enhancement on a picture taken at a low-lighting condition. The original picture is the one in the center. Note the accentuated sensor noise, typical of the low-light condition. The picture on the left is the result after independently processing each color. Finally, the picture on the left is provided for comparison. It is a wavelet-based de-noising, with a soft threshold [8].

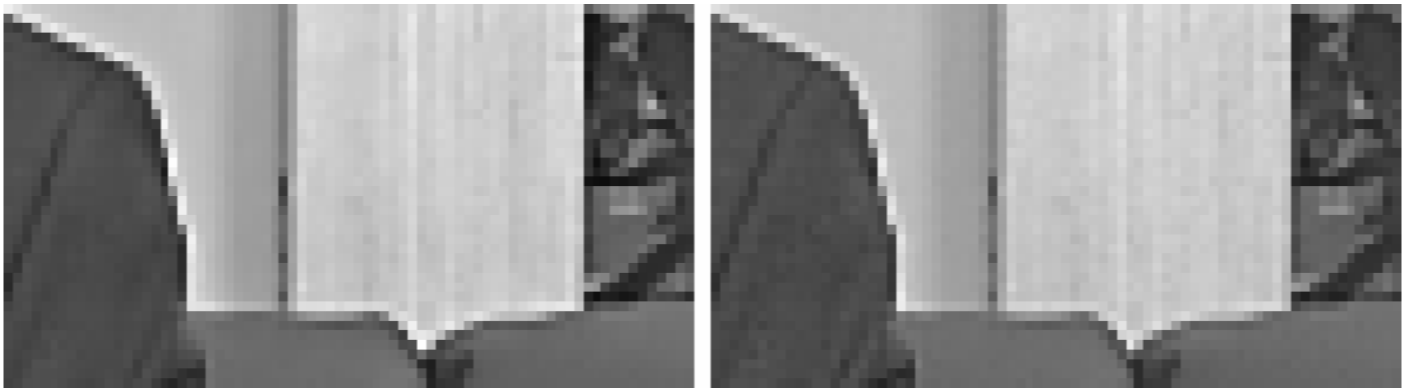

Figure 7 – *Detail on Salesman: Processed (left) and Original (right)*

Figure 8 – *A color picture on low-lighting. Original (center), Proposed (left) and wavelet de-noising (right).*

V. Conclusions

We have presented an algorithm for reducing camera (sensor) noise in image sensors. The algorithm can differentiate between camera noise, and the actual frequency content of the image. The algorithm does that by using an auxiliary picture, where the sensor noise is independent, but picture content is assumed to be the same. Artifacts were avoided by limiting the filtering to spatial domain, in contrast to other techniques that do median or some other type of time-domain filtering. The basic operation of the algorithm is simple, and should be able to be handled by the camera. We tested the algorithm on frames obtained from video sequences, as well as frames obtained from a still camera. In the experiments with the camera, we set the camera for under-exposure, as would be the case in low-light conditions. We then used the algorithm independently on each color component. The reduction in the granularity is easily observed, and the results compare well to wavelet based soft thresholding, and to other techniques we have considered.

References

[1] F. Dekeyser, P. Bouthemy, and P. Perez, "Spatial-temporal Wiener filtering of image sequences using a parametric model," *Proc. ICIP'00*, pp. 1586-89, 2000.

[2] P. R. Giaconni, G. A. James, S. Minelly, and A. Curley, "Motion-compensated multi-channel noise reduction of colour film sequences," *SPIE Journal of Electronic Imaging*, vol. 8-3, pp. 246-254, July 1999.

[3] K. J. Boo and N. K. Bose, "A motion-compensated spatial-temporal filter for image sequences with signal dependent noise," *IEEE Trans. Cir. Syst. For Video Tech.*, vol. 8-3, pp. 287-298, June 1998.

[4] A. Kundu, "Motion estimation by image content matching and application to video processing," *Proc. ICASSP'96*, vol.IV, pp. 1902-5, 1996.

[5] G. Perschigg, et all, "Digital Photography with flash and non-flash pairs," *Proc. of SIGGRAPH, 2004*.

[6] H. Cheong, A Tourapis, J. Llach, and J. Boyce, "Adaptivespatio-temporal filtering for video de-noising," *Proc. of ICIP*, 2004.

[7] D. Florencio, "Motion sensitive pre-processing for video," *Proc. of ICIP*, 2001.

[8] Donoho, D.L., "De-noising by soft-thresholding," *IEEE Trans. on Inf. Theory*, 41, 3, pp. 613-627, May 1995.