Image Sequences Noise Reduction: An Optical Flow Based Approach *

Roman Dudek, Carmelo Cuenca, and Francisca Quintana

Departamento de Informática y Sistemas,
Campus Universitario de Tafira, Las Palmas GC, Spain
roman@idecnet.com, {ccuenca,fquintana}@dis.ulpgc.es

Abstract. We present an optical flow based method for noise reduction in image sequences. To prevent artefacts caused by optical flow imperfections, we propose a method to estimate these imperfections. We use the estimation to adaptively choose either a temporal or a spatial based noise reduction algorithm to be applied in different image zones. Our results have shown that an important noise reduction can be achieved with the proposed method, without the drawbacks of the simpler methods. The method has provided important noise reductions even with complex image sequences.

1 Noise in Film

Either we use digital or chemical resources for shooting image sequences, an important amount of noise is introduced due to the physical nature of the shooting process. This noise needs to be frequently reduced to make further digital processing of the film image sequences. Although many people consider that they miss something important if noise is removed from the grainy film look of the movie theater projections, noise is a technical problem that must be reduced because of their interferences with the digital postproduction process. The noise can be easily added after the image processing, if desired so, for artistic reasons. Multiple methods were proposed and used for the noise reduction.

2 Usual Methods of Noise Reduction and Our Proposed Method

The most obvious approach to reduce image noise is by suppressing the high frequencies in the image spectrum characteristic, for example with Gaussian filtering. While this technique reduces the image noise, also some important details of the image would be removed. Adaptive methods have also been developed, aimed at distinguishing for example the image edges and do not blur across them. However, every such method has its limitations: For example, the edge

* This work has been partially supported by The Spanish Ministry of Science and Innovation under contract TIN2007-60625 and by FEDER funds.

detecting method would blur any detailed textures with not enough contrast to be classified as edges. Another approach is based on averaging multiple adjacent images of an image sequence. However, this method only works well for static scenes as any moving object in the scene creates ghost-like trails.

In our work, we propose and test a noise reduction method based on averaging the image pixel values along the sequence time. As in the simple frame averaging method described above, we will use the sequence time consistency to distinguish noise from actual image. However, to extend the method's applicability to moving objects, we will use the optical flow (OF) to track each pixel's actual position in previous and following images.

3 Noise Reduction Based on Image Averaging and Estimation of the Noise Reduction

The noise of each image pixel is caused by the accumulation of different phenomena, like thermal noise and random particle distribution[1]. All these phenomena usually have a Gaussian distribution, so we can consider that the noise itself would have a Gaussian distribution. Experimental data confirms this assumption.

Let us suppose the value of a given pixel belonging to the image I_n to be the sum of an ideal signal S_n and a noise N_n with Gaussian distribution and zero mean value:

$$I_n = S_n + N_n(0, \sigma^2)$$

As previously mentioned, there is no relation between noise patterns in consecutive frames. The noise distribution for a set of pixels, each one belonging to a different frame, can be considered purely random. On the other hand, the signal value S should be constant in case of static scenes, or in cases of dynamic scenes if we account for movements using OF information, as we will show later.

In our proposed method, we calculate the average for each pixel across n frames,

$$I_{average} = \frac{\sum_{i=1}^{n} S + N_i(0, \sigma^2)}{n} = S + \frac{\sum_{i=1}^{n} N_i(0, \sigma^2)}{n}$$

We can show that while the signal value S is preserved, the dispersion of the averaged noise amplitude is reduced by a factor of $\sqrt{n}$, the number of summed values. However, as the noise energy is proportional to the square of the amplitude, the noise energy is reduced by a factor of n.

4 Optical Flow Based Noise Reduction

In our process, we first calculate the OF fields[2][3][4] for each pair of consecutive images of the sequence, for both forward and backward directions of the OF. Once the OF fields are calculated, we perform the actual image averaging:

for each pixel of a given image, we track the corresponding pixel coordinates in certain number of images preceding and following the current image in the sequence[5][6]. The coordinates are calculated recursively, following the track of the pixel along the sequence.

Notice that only the (x_0, y_0) coordinates are integer values. The (x_n, y_n) pairs are typically non-integer, so interpolation is used in order to obtain the $u_n(x_n, y_n)$ and $v_n(x_n, y_n)$ values.

Then, the resulting pixel value is calculated as the average value of the tracked pixels:

$$I_{average}(x, y) = \frac{\sum_{n=-r}^{r} I(x_n, y_n)}{2r + 1} \qquad (1)$$

Again, interpolation is used to obtain the image values for non-integer coordinates (x_n, y_n).

Selecting the value of r (number of frames we use for averaging) requires a compromise as increasing its value increases the grade of noise reduction but it also increases the demands on the OF precision, as inaccuracies in the fields would accumulate while tracking individual pixels along more frames. Our tests show that the best results are obtained by averaging 3 to 7 frames, i.e., $r \in [1, 3]$.

4.1 Treating Zones of Large Errors in OF

The main limitations of the method are the precision of the OF fields and the occlusions or large changes in illumination in the image sequence, which can prevent us from following a given pixel position over long sequences of images.

As a first measure to minimize these problems, during the calculation of the noise reduced current frame, we use a group of images consisting of both the previous and the following frames, with the current frame in the center of the group. This leads to a lower error accumulation than, for example, only using frames which are previous to the current one.

Even with perfect OF fields available, the time averaging method will fail in zones of occlusions. Practically, the OF obtained by an estimation method will contain certain errors, with problematic zones containing large errors. Experiments show that ignoring such errors can cause important artifacts in zones of the scene where the OF field is incorrect (or even undefined in case of occlusions or transparencies). We developed a method to detect such zones and threat them differently.

4.2 Detecting Zones of Large Errors in OF

A good detection of the OF validity is needed in order to distinguish where OF based method is applicable and where intra-frame filtering should be used instead.

We can assume that where the OF vectors got "lost", and could not correctly follow the movements in the scene, the values of pixels that we find in the neighboring frames using these OF vectors will differ. To estimate these OF errors for each result pixel, we propose to calculate the dispersion (medium square error) of the values used for the calculation of the average value of each result pixel:

$$E(x,y) = \frac{\sum_{n=-r}^{+r} \left(I_n(x_n,y_n) - I_{average}(x,y)\right)^2}{2r+1} \tag{2}$$

Unfortunately, considering that the number of averaged values in the above sums is relatively low (3 to 7), the error measure $E(x,y)$ itself contains a large amount of noise.

To analyze the properties of $E(x,y)$, we can express it as a sum of the "OF Error" $E_{OF}(x,y)$, caused only by the OF imperfections, and a noise component $N_{error}(x,y)$:

$$E(x,y) = E_{OF}(x,y) + N_{error}(x,y) \tag{3}$$

The amplitude of $N_{error}(x,y)$ can be derived from the amplitude of the noise in the source images. Let the individual source image pixels be considered the sum of the "Signal" and a "Noise":

$$I_n(x,y) = S_n(x,y) + N_n(x,y) \tag{4}$$

Let us suppose that the OF vectors were perfect, so the $S_n(x,y)$ are equal for each n from $[-r,+r]$, and $E_{OF}(x,y) = 0$. Then,

$$N_{error}(x,y) = E(x,y) = \frac{\sum_{n=-r}^{r} \left(I_n(x_n,y_n) - I_{average}(x,y)\right)^2}{2r+1} \tag{5}$$

The value $N_{error}(x,y)$ would converge to a certain constant c for large values of r, with a high number of averaged frames. This constant could be used as a threshold to decide the validity of the OF fields for a certain pixel: If the error $E(x,y)$ is similar to c, the pixel was likely correctly followed. If the $E_{OF}(x,y)$ is much higher than c, the OF is likely invalid for this pixel, and time averaging should not be used. We use simple spatial Gaussian filtering of the image for the given pixel instead:

$$I_{result} = \begin{cases} (g \circ I_0)(x,y)) \text{ for} E(x,y) > c \\ I_{average}(x,y) \text{ for} E(x,y) \leq c \end{cases} \tag{6}$$

However, as we need to use only low values of r ($r \in [1,3]$), the $E(x,y)$ values do not converge enough to the actual noise level of the sequence. There is an important noise component in our $E(x,y)$ itself, making impractical such direct decision per pixel. In practice, the random dispersion of the total error measure $E(x,y)$ is frequently larger than the $E_{OF}(x,y)$ we are actually trying to detect, so neighboring pixels would be frequently randomly misclassified due to this noise component.

To make a better classification, we need to reduce the randomness in the $E(x,y)$ error measure field. We propose to carry out a spatial averaging of the

$E(x, y)$ field. This spatial averaging, added to the temporal averaging used to create the $E(x, y)$ itself, can largely reduce the randomness of the $E(x, y)$ error measure, while not affecting much the $E_{OF}(x, y)$ component that we are trying to detect, supposing that it is locally smooth anyway. We used Gaussian filter for this averaging operation, to create a filtered error measure $E'(x, y)$

$$E'(x, y) = (g \circ E)(x, y)) \tag{7}$$

The modified algorithm to obtain the final result will use $E'(x, y)$ instead of $E(x, y)$:

$$I_{result} = \begin{cases} (g \circ I_0)(x, y)) \text{ for} E'(x, y) > c \\ I_{average}(x, y) \text{ for} E'(x, y) \leq c \end{cases} \tag{8}$$

Using $E'(x, y)$, our threshold classification method provides a much better detection of the problematic zones. However, the application of a thereshold classification causes some visible artifacts along the edges of the zones where the decision changes. In order to reduce such artifacts, we propose to create a thin transition zone using a clamped blending equation instead of a thresholding:

$$p = clamp((E'(x, y) - c) * s) \tag{9}$$

where s is a user defined constant and $clamp()$ is a function limiting p to the range $[0, 1]$. Then, the final result is obtained by using a blending equation instead of using a threshold:

$$I_{result}(x, y) = p * (g \circ I_0)(x, y)) + (1 - p) * I_{average}(x, y) \tag{10}$$

In our test application, the constant s is a user defined value adjusted in such a way that a transition zone of only a few pixels wide will be created around OF error zones. Similarly, c is adjusted by the user in order to correctly detect the zones of OF errors, while ignoring errors too small to cause visible artifacts in the result. The adjustment of other parameters, like the Gaussian filtering radius, depend on the resolution and noise level of the used image sequence. However, once these values are adjusted, they seem to be constant for all shots from the same original negative film roll, so only few adjustments need to be done for each film project.

5 Results

For our tests we used a variety of digitalized image sequences, originally shot on negative film. We obtained important noise reductions without observably suppressing any detail in the scene.

Figure 1 shows a sample image from the original image sequence. This image is a part of a sequence filmed intentionally on $8mm$ celluloid film, to obtain obvious film look even when broadcast over standard PAL television. We have chosen this material in order to to make observable the necessary details in a

Fig. 1. Sample image from the original sequence.

Fig. 2. Sample image after simple frame averaging.

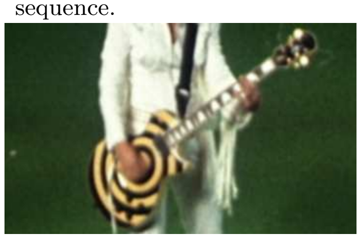

Fig. 3. Sample image after applying a simple Gaussian filter.

Fig. 4. Sample image after averaging using OF based, with pixel following over 5 frames.

printed form of this document. We can observe important noise caused by film grain, specially in the flat background behind the musician.

In Figure 2 we can see the result obtained after a simple frame averaging of five frames. We can see that the image detail is lost due to the motion trails created around any object in movement. These trails can be clearly observed on the right elbow of the musician. The noise level is generally reduced, as can be obviously observed in the flat background, for example. However image stays sharp only in the few parts where there was no movement in the range of the five frames used.

After applying a simple Gaussian filtering to the image we obtain the Figure 3. The Gaussian filter radius was set just large enough to provide a visually similar noise reduction as it would be obtained by averaging five frames of the sequence. We can see that there are no trails around the shoulder as in Figure 2, but edges are visibly blurred. Compare the black shoulder belt edges, for example. Some image detail is visibly lost.

Figure 4 shows the frame averaging of 5 frames using OF based tracking of the pixel positions. We can look at that the noise is reduced. Neither there are no trails of simple frame averaging, nor there is the detail loss of the spatial Gaussian filtering. However, artifacts can be found in some zones. For example, see the hand hitting the strings of the guitar: some black strips, not present in the original image, can be observed over the hand. This type of artefact is characteristic for the zones where the OF filed calculation failed due to some reason, because of a large advancing occlusion in this case.

Fig. 5. Sample image using our adaptive algorithm. A simple, hard threshold decision is used in this case.

Fig. 6. Sample image using our adaptive algorithm. Improved threshold is used.

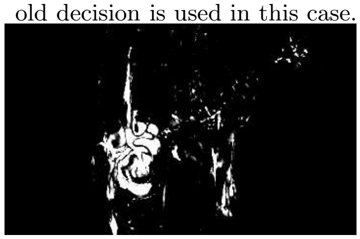

Fig. 7. Mask image resulting from a simple threshold applied to detect errors in OF.

Fig. 8. Mask image resulting from the improved threshold, using a filtered error measure and smooth transitions.

Figure 5 is the result of our adaptive algorithm based on the detection of large errors in the OF. The OF corrected frame averaging is used where errors are small. Gaussian filtering is used where OF errors are large. Virtually no artefacts can be observed here, however the hard threshold decision based on unfiltered error measure can be observable in the form of visible spots in some scenes. In this case, a spot is hardly observable on the fingers of the hand hitting the strings.

In Figure 6 improved threshold is used, with filtered error measure and smooth transition between the zones. It is very difficult to find any visible artefacts anymore, anywhere in the image. Some parts of the image, like the hand hitting the strings, may look blurry. However, comparing with the unfiltered original image, and you will find that the hand was blurry in the original too, due to non-zero exposure time of the frame. The resulting image preserves, even makes clearer, any detail present in the original sequence, while reducing the noise in the same time.

Figure 7 shows the mask resulting from a simple threshold applied to detect errors in OF. White pixels represent the zones of large error, where Gaussian filtering image will be used. Black pixels represent the zones of apparently correct OF, where the preferable OF corrected time averaging will be used. Observe the zone of the hand striking the strings: While almost the whole zone is white, detected as containing large errors, there are some unexpected, spurious strips of black pixels. These false negatives (zones of error detected as correct) cause the spots described above.

Figure 8 shows the mask resulting from improved threshold method, using filtered error measure and smooth transition. We can see that the spurious black strips in the zone of the moving hand were eliminated due to the error measure filtering. Softening the outlines, using a soft threshold instead of a hard one, further reduces the visibility of any spot artefacts left.

6 Conclusions and Future Work

In this work we have presented an adaptive optical flow based noise reduction method, combining the advantages of both spatial and temporal filtering methods. We proposed and tested a robust tresholding method to detect flaws in the OF, an choose the prefferable method for each image zone, with good transitioning on the borders between the zones. We have shown that an important noise reduction can be achieved with the proposed method, without the drawbacks of the simpler methods. While the method requires some manual parameter adjusting, virtually in all practical test cases we were able to achieve visually artefact-less noise reduction.

In our future work, we will try to detect different cases of problematic zones, and test alternative methods for improving the results in such zones. Occlusions could be detected and handled explicitly in order to stop tracking a pixel in images where that pixel is not present anymore, while still using lower number of samples for averaging. This could be combined with a smaller amount of spatial filtering to complement the noise reduction already achieved by partial frame averaging, instead of doing a complete fallback to spatial filtering.

Also the natural blurring in zones of movement is anamorphic: There is a loss of resolution in direction of the movement, but not in the direction perpendicular to it. As a consequence, using an anamorphic filtering instead of symmetric Gaussian filter would likely improve the results.

References

1. Minelly, S. Curley, A. Giaccone, P. and Jones, G.A. Reducing chromatic grain noise in film sequences (1998)
2. Álvarez, L., Weickert, J., and Sánchez J., Reliable Estimation of Dense Optical Flow Fields with Large Displacements, *International Journal of Computer Vision*,39(1), 41–56, 2000
3. S. S. Beauchemin and J. L. Barron, The Computation of Optical Flow, *ACM Computing Surveys*, volume 27(3), 433–467, 1995
4. Horn B., and Schunk B., Determining Optical Flow, *AI Memo 572*, Massachusetts Institue of Technology, 1980
5. Zhongding Jiang, Tien-Tsin Wong and Hujun Bao. Practical Super-Resolution from Dynamic Video Sequenes. In Proceedings of IEEE Computer Vision and Pattern Recognition 2003 (CVPR 2003), Madison, Wisconsin, USA, June 16-22, 2003.
6. Sean Borman and Robert L. Stevenson, Spatial Resolution Enhancement of Low-Resolution Image Sequences: A Comprehensive Review with Directions for Future Research, Department of Electrical Engineering, University of Notre Dame, 1998