

Adaptive enhancement and noise reduction in very low light-level video

Henrik Malm Magnus Oskarsson Eric Warrant
Lund Vision Group, Department of Cell and Organism Biology
Lund University, Helgonavägen 3, S-223 62 Lund, Sweden
henrik.malm@cob.lu.se

Petrik Clarberg Jon Hasselgren Calle Lejdfors
Department of Computer Science, Lund Institute of Technology
John Ericssons väg, 223 00 Lund, Sweden

Abstract

A general methodology for noise reduction and contrast enhancement in very noisy image data with low dynamic range is presented. Video footage recorded in very dim light is especially targeted. Smoothing kernels that automatically adapt to the local spatio-temporal intensity structure in the image sequences are constructed in order to preserve and enhance fine spatial detail and prevent motion blur. In color image data, the chromaticity is restored and demosaicing of raw RGB input data is performed simultaneously with the noise reduction. The method is very general, contains few user-defined parameters and has been developed for efficient parallel computation using a GPU. The technique has been applied to image sequences with various degrees of darkness and noise levels, and results from some of these tests, and comparisons to other methods, are presented. The present work has been inspired by research on vision in nocturnal animals, particularly the spatial and temporal visual summation that allows these animals to see in dim light.

1. Introduction

1.1. Biological Inspiration

In order to survive in habitats with high competition for resources and the threat of predation, many animals have evolved a nocturnal lifestyle. To be able to navigate and forage during the hours of darkness, these animals have developed excellent vision at light levels where humans are practically blind, even though the optical properties of their eyes are in many cases inferior to those found in the eyes of humans. It is therefore interesting to study the visual systems of these animals to gain insight into the mechanisms that permit these impressive visual capabilities.

A characteristic of visual systems in nocturnal animals is the highly developed ability to sum the visual signal locally in space and time to increase signal strength and improve the reliability of intensity estimations. This summation mechanism is adaptive and becomes more pronounced at lower light levels, but comes at the cost of lower spatial and temporal resolution. The summation can either take place at the receptor level or higher up in the visual pathway.

The relative amount of spatial and temporal summation depends on the specific needs and lifestyles of the animal in question. Nocturnal animals that move fast and need to react to high speed events, *e.g.* flying moths, require good temporal resolution, therefore favoring spatial summation over temporal summation. In contrast, in nocturnal animals that move slowly temporal summation is favoured in order to maximize spatial resolution. Warrant [25] constructed a model for visual signal processing in animals that can be used to find the optimal extents of spatial and temporal summation that maximize visual performance at a given light intensity and image velocity. This biological mechanism of adaptive spatio-temporal smoothing was the initial inspiration for the work presented in this paper.

1.2. Image Processing Objective

Modern digital cameras still rely on a single exposure time and on imaging sensors that possess photo elements of uniform sensitivity. This often results in underexposed dark areas in the obtained images. The problem is especially pronounced in video cameras where the exposure time is limited to prevent motion blur. Most people have probably experienced this problem when shooting a movie either indoors or within a scene where the intensities vary by several orders of magnitude.

If we simply try to amplify the dark areas, uniformly or by a non-linear intensity transformation (*tone mapping*), the

low signal-to-noise ratio (SNR) in these areas will become very apparent. There are a number of sources of noise in a camera, *c.f.* [19, 1], including photon shot noise and readout noise, which together constitute the main problems. In this work we aim to enhance very dark image sequences that have an extremely low SNR.

To reduce the noise significantly we apply adaptive spatio-temporal intensity smoothing. For every pixel in the image sequence a specific summation kernel is constructed which adapts to the local spatio-temporal intensity structure. This kernel is then used to calculate a weighted mean in the neighborhood of the current pixel. Instead of applying a combination of pure spatial and pure temporal summation, as in the model of Warrant, we view the three dimensional spatio-temporal space symmetrically and shape and direct the kernels according to 3D intensity structure. By doing this, we find an optimal way of smoothing the intensities, while preserving, and even enhancing, important object contours and preventing motion blur.

After the noise has been efficiently reduced, the intensities are transformed using tone mapping, implemented by a contrast limited version of histogram equalization, in order to spread out the intensities more uniformly and widen the dynamic range of the darker areas.

The algorithm works very well on color image data and maximally restores the original chromaticity values as seen in brighter lighting, along with the luminance. When dealing with raw color image data obtained in a Bayer pattern, the interpolation to a three-channel RGB image is done concurrently with the spatio-temporal smoothing.

The implementation of the algorithm is quite simple and is very well suited for parallel processing on a graphic processing unit (GPU). In this way the complete processing can, at time of writing, be performed interactively, with approx. 7 fps, on a 360x288 pixels color image sequence.

2. Related Work

There exist a multitude of noise reduction techniques that apply spatio-temporal weighted averaging for noise reduction purposes. Many authors have additionally realized the benefit of trying to reduce the noise by filtering the sequences along motion trajectories in spatio-temporal space and in this way, ideally, avoid motion blur and unnecessary amounts of spatial averaging. These noise reduction techniques are usually referred to as motion-compensated (spatio-)temporal filtering. In [11], means along the motion trajectories are calculated while in [29] and [16] weighted averages, dependent on the intensity structure and noise in a small neighbourhood, are applied. In [21], so-called linear minimum mean square error filtering is used along the trajectories. Most of the motion-compensated methods use some kind of block-matching technique for the motion estimation that usually, for efficiency, employs rather few im-

ages, commonly 2 or 3. However, in [21], [29] and [6] variants using optical flow estimators are presented. Drawbacks of the motion-compensating methods is that the filtering relies on a good motion estimation to give a good output, without excessive blurring. The motion estimation is especially complicated for sequences severely degraded by noise. Different approaches have been applied to deal with this problem, often by simply reducing the influence of the temporal filter in difficult areas. This, however, often leads to disturbing noise at for example object contours.

An interesting family of smoothing techniques for noise reduction are the ones that solve an edge-preserving anisotropic diffusion equation on the images. This approach was pioneered by Perona and Malik [17] and has had many successors, including the work by Weickert [26]. These techniques have also been extended to 3D and spatio-temporal noise reduction in video, *c.f.* [24] and [13]. In [24] and [27], the so-called *structure tensor* or *second moment matrix* is applied in a similar manner to our approach in order to analyze the local spatio-temporal intensity structure and steer the smoothing accordingly. The drawbacks of techniques based on diffusion equations include the fact that the solution has to be found using an often time-consuming iterative procedure. Moreover, it is very difficult to find a suitable stopping time for the diffusion, at least in a general and automatic manner. These drawbacks make these approaches in many cases unsuitable for video processing.

A better approach is to apply single-step structure-sensitive adaptive smoothing kernels. The bilateral filters introduced by Tomasi and Manduchi [23] for 2D images falls within this class. Here, edges are maintained by calculating a weighted average at every point using a Gaussian kernel, where the coefficients in the kernel are attenuated based on how different the intensities are in the corresponding pixels compared to the center pixel. This makes the local smoothing very dependent on the correctness in intensity in the center pixel, which cannot be assumed in images heavily disturbed by noise.

An approach that is closely connected to both bilateral filtering, and to anisotropic diffusion techniques based on the structure tensor, is the structure-adaptive anisotropic filtering presented by Yang *et al.* [28]. In our search for an optimal strategy to apply spatio-temporal summation for noise reduction in low light-level video, we naturally settled for this approach. Since our current methodology is an extension of this technique we will present it in detail in Section 3. For a study of the connection between anisotropic diffusion, adaptive smoothing and bilateral filtering, see [2].

Another group of algorithms connected to our work are in the field of high dynamic range (HDR) imaging, *c.f.* [12, 20] and the references therein. The aim of HDR imaging is to alleviate the restriction caused by the low dynamic range of ordinary CCD cameras, *i.e.* the restriction to the

ordinary 256 intensity levels for each color channel. Most HDR imaging algorithms are based on using multiple exposures of a scene with different settings for each exposure and then using different approaches for storing and displaying this extended image data. However, the HDR techniques are not especially aimed at the kind of low-light level data which we are targeting in this work, where the utilized dynamic range in the input data is in the order of 5 to 20 intensity levels and the SNR is extremely low.

Surprisingly few published studies exist that especially target noise reduction in low light-level video. Three recent examples are the above-mentioned method by Bennett and McMillan [3], the technique presented by Lee *et al.* [14] and the approach by Malm and Warrant [15]. In [14], very simple operations are combined in a system presumably developed for easy hardware implementation in *e.g.* mobile phone cameras and other compact digital video cameras. In our tests of this method, which we comment on in Section 6.1, it is evident that this method cannot handle the high levels of noise which we target in this work.

The approach taken by Bennett and McMillan [3] for low dynamic range image sequences is more closely connected to our technique. Their virtual exposures framework includes the bilateral ASTA-filter (Adaptive Spatio-Temporal Accumulation) and a tone mapping technique. The ASTA-filter, which changes to relatively more spatial filtering in favour of temporal filtering when motion is detected, is in this way related to the biological model of Warrant [25]. However, since bilateral filters are applied, the filtering is edge-sensitive and the temporal bilateral filter is additionally used for the local motion detection. The filters apply novel dissimilarity measures to deal with the noise sensitivity of the original bilateral filter formulation. The filter size applied at each point is decided by an initial calculation of a suitable amount of amplification using tone mapping of an isotropically smoothed version of the image. A drawback of the ASTA-filter is that it requires global motion detection as a pre-processing step to be able to deal with moving camera sequences. The sequence is then warped to simulate a static camera sequence. The succeeding tone mapping approach proposed by Bennett and McMillan is further discussed in Section 4. We have implemented the complete virtual exposures framework and will compare our proposed method to this approach in Section 6.1.

In [15], we observed the usefulness of applying the structure tensor for single-step adaptive spatio-temporal smoothing for noise reduction in low light-level video. However, the technique presented there requires a pre-processing step consisting of foreground-background segmentation in the style of Stauffer and Grimson [22] for static camera sequences. Further, similar to the approach of Bennett and McMillan, it requires an ego-motion estimation to deal with moving camera sequences.

In this paper we generalize the ideas from [15] so that the adaptive spatio-temporal smoothing is applied in a uniform way for both static and moving sequences, without the need for any pre-processing step. This simple structure makes our approach ideal for efficient parallel implementation on a GPU. Additionally, an approach for sharpening of the most important features is suggested to prevent over-smoothing. In contrast to [15], we here also target color input data. Raw color input data is interpolated from the Bayer pattern simultaneously to the adaptive smoothing.

3. Adaptive Spatio-Temporal Smoothing

We will now review the structure-adaptive anisotropic image filtering by Yang *et al.* [28] and present our modifications and extensions to improve its applicability and make it suitable for our low-light level vision objective.

The method computes a new image $f_{out}(x)$, by applying at each spatiotemporal point $x_0 = (x'_0, y'_0, t'_0)$, a kernel $k(x_0, x)$ to the original image $f_{in}(x)$ such that:

$$f_{out}(x_0) = \frac{1}{\mu(x_0)} \int \int \int_{\Omega} k(x_0, x) f_{in}(x) dx, \quad (1)$$

where

$$\mu(x_0) = \int \int \int_{\Omega} k(x_0, x) dx \quad (2)$$

is a normalizing factor. The normalization makes the sum of the kernel elements equal to 1 in all cases, so that the mean image intensity doesn't change. The area Ω over which the integration, or in the discrete case, summation, is made is chosen as a finite neighborhood centered around x_0 .

Since we want to adapt the filtering to the spatiotemporal intensity structure at each point, in order to reduce blurring over spatial and temporal edges, we calculate a kernel $k(x_0, x)$ individually for each point x_0 . The kernels should be wide in directions of homogeneous intensity and narrow in directions with important structural edges. To find these directions, the intensity structure is analyzed by the so-called *structure tensor* or *second moment matrix*. This object has been developed and applied in image analysis in numerous papers *e.g.* by Jähne and co-workers [10]. The tensor $J_{\rho}(x_0)$ is defined in the following way:

$$J_{\rho}(\nabla f(x_0)) = G_{\rho} \star (\nabla f(x_0) \nabla f(x_0)^T), \quad (3)$$

where

$$\nabla f(x_0) = \left[\frac{\partial f}{\partial x'_0} \frac{\partial f}{\partial y'_0} \frac{\partial f}{\partial t'_0} \right]^T \quad (4)$$

is the spatiotemporal intensity gradient of f at the point x_0 . G_{ρ} is the Gaussian kernel function

$$G_{\rho}(x) = \frac{1}{\mu} e^{-\frac{1}{2} \left(\frac{x'^2 + y'^2 + t'^2}{\rho^2} \right)}, \quad (5)$$

where μ is the normalizing factor. The notation \star means elementwise convolution of the matrix $\nabla f(x_0)\nabla f(x_0)^T$ in a neighborhood centered at x_0 . It is this convolution that gives us the smoothing in the direction of gradients which is the key to the noise insensitivity of this method.

Eigenvalue analysis of J_ρ will now give us the structural information that we seek. The eigenvector v_1 corresponding to the smallest eigenvalue λ_1 will be approximately parallel to the direction of minimum intensity variation while the other two eigenvectors are orthogonal to this direction. The magnitude of each eigenvalue will be a measure of the amount of intensity variation in the direction of the corresponding eigenvector. For a deeper discussion on eigenvalue analysis of the structure tensor see [8].

The basic form of the kernels $k(x_0, x)$ that are constructed at each point x_0 is that of a Gaussian function,

$$k(x_0, x) = e^{-\frac{1}{2}(x-x_0)^T R \Sigma^2 R^T (x-x_0)} , \quad (6)$$

including a rotation matrix R and a scaling matrix Σ . The rotation matrix is constructed from the eigenvectors v_i of J_ρ ,

$$R = [v_1 \quad v_2 \quad v_3] , \quad (7)$$

while the scaling matrix has the following form,

$$\Sigma = \begin{bmatrix} \frac{1}{\sigma(\lambda_1)} & 0 & 0 \\ 0 & \frac{1}{\sigma(\lambda_2)} & 0 \\ 0 & 0 & \frac{1}{\sigma(\lambda_3)} \end{bmatrix} . \quad (8)$$

The function $\sigma(\lambda_i)$ is a decreasing function that sets the width of the kernel along each eigenvalue direction. The theory in [28] is mainly developed for 2D images and measures of corner strength and of anisotropism, both involving ratios of the maximum and minimum eigenvalues, are there calculated at every point x_0 . An extension of this to the 3D case is then discussed. However, we have not found these two measures to be adequate for the 3D case since they give too much focus on singular corner points in the video input and to a large extent disregards the linear and planar structures that we want to preserve in the spatiotemporal space. For example, a dependence of the kernel width in the temporal direction on the eigenvalues corresponding to the spatial directions does not seem appropriate in a static background area. We instead simply let an exponential function depend directly on the eigenvalue λ_i in the current eigenvector direction v_i in the following way,

$$\sigma(\lambda_i, x_0) = \begin{cases} \Delta\sigma e^{-\frac{\lambda_i}{d} + \frac{2}{5}} + \sigma_{min}, & \lambda_i > 2d/5 \\ \sigma_{max}, & \lambda_i \leq 2d/5 \end{cases} \quad (9)$$

where $\Delta\sigma = (\sigma_{max} - \sigma_{min})$, so that σ_i attains its maximum σ_{max} below $\lambda = 2d/5$ and asymptotically approaches its minimum σ_{min} when $\lambda \rightarrow \infty$. The parameter d scales

the width function along the λ -axis and has to be set in relation to the current noise level. Since the part of the noise that stems from the quantum nature of light, *i.e.* the photon shot noise, depends on the brightness level, it is signal-dependent and the parameter d should ideally be set locally. Future research will be focused on an adaptive strategy for setting this parameter automatically. However, it has been noticed that when the type of camera and the type of tone mapping is fixed, a fixed value on d usually works for a large part of the dynamic range. When changing the camera and tone mapping approach, a new value of d has to be found for optimal performance.

When the widths $\sigma(\lambda_i, x_0)$ have been calculated and the kernel subsequently constructed according to 6, equation 1 is used to calculate the output intensity f_{out} of the smoothing stage in the current pixel x_0 .

3.1. Considerations for Color

The discussion so far has dealt with intensity images. We will here discuss some special aspects of the algorithm when it comes to processing color images.

In applying the algorithm to RGB color image data one could envision a procedure where the color data in the images are first transformed to another color space including an intensity channel, *e.g.* the HSV colour space, *c.f.* [7]. The algorithm could then be applied unaltered to the intensity channel, while smoothing of the other two channels could either be performed with the same kernel as in the intensity channel or by isotropic smoothing. The HSV image would then be transformed back to the RGB color space.

However, in very dark color video sequences there is often a significant difference in the noise levels in the different input channels: for example the blue channel often has a relatively higher noise level. It is therefore essential that there is a possibility to adapt the algorithm to this difference. To this end, we chose to calculate the structure tensor J_ρ , and its eigenvectors and eigenvalues, in the intensity channel, which we simply define as the mean of the three color channels. The widths of the kernels are then adjusted separately for each colour channel by using a different value of the scaling parameter d for each channel. This gives a clear improvement of the output with colors that are closer to the true chromaticity values and with less false color fluctuations than in the above mentioned HSV approach.

When acquiring raw image data from a CCD or CMOS sensor, the pixels are usually arranged according to the so-called Bayer pattern. It has been shown, *c.f.* [9], that it is efficient and suitable to perform the interpolation from the Bayer pattern to three separate channels, so-called *demosaicing*, simultaneously to the denoising of the image data. We apply this approach here, for each channel, by setting to zero the coefficients in the summation kernel $k(x_0, x)$ corresponding to pixels where the intensity data isn't avail-

able and then normalizing the kernel. A smoothed output is then calculated for both the noisy input pixels and the pixels where data are missing.

3.2. Sharpening

The structure tensor is surprisingly good at finding the direction of motion in very noisy input sequences. However, in some cases the elongation of the constructed summation kernels can be slightly misaligned with the motion direction. In these cases the contours of the moving objects can be somewhat blurred. As an alternative to our standard adaptive smoothing approach we propose the addition of some high-boost filtering to sharpen up these contours. The high-boost filter is defined as $3 \times 3 \times 3$ tensor with the value -1 in all elements except the center element which has the value $27A - 1$, *c.f.* [7]. We have attained the best results using $A = 1.2$. If the filter is applied after the complete smoothing process we will run into problems of negative values and rescaling of intensities, which will give a global intensity flickering in the image sequence. We instead propose that, at each point x_0 , the constructed smoothing kernel $k(x_0, x)$ is filtered with the high-boost filter yielding a new kernel \hat{k} . The kernel is then normalized so that its coefficients add up to 1. In this way, application of the constructed kernels $\hat{k}(x_0, x)$ will add some sharpening at the most important spatial and temporal edges together with the general adaptive smoothing.

4. Tone Mapping

The adaptive smoothing of the input data presented above reduces the noise in the low-light level data, but the images are still just as dark as in the unprocessed input data. The next step is to apply an amplifying intensity transformation such that the dynamic range of the dark areas is increased while preserving the structure in brighter areas, if such areas exist. The procedure of intensity transformation is also commonly referred to as *tone mapping*. The tone mapping could actually be performed either before or after the noise reduction, with a similar output, as long as the scaling parameter d for the smoothing kernels is chosen to fit the current noise level.

In the virtual exposures method of Bennett and McMillan [3], a tone mapping procedure is applied where the actual mapping is a logarithmic function similar to the one proposed by Drago *et al.* [5]. The tone mapping procedure also contains additional smoothing using spatial and temporal bilateral filters and an attenuation of details, found by the subtraction of a filtered image from a non-filtered image. We instead choose to do all smoothing in the separate noise reduction stage and here concentrate on the tone-mapping.

The tone mapping procedure of Bennett involves several parameters, both for the bilateral smoothing filters and for

changing the acuteness of two different mapping functions, one for the large scale data and one for the attenuation of the details. These parameters have to be set manually and will not adapt if the lighting conditions change in the image sequence. Since we aim for an automatic procedure we instead opt for a modified version of the well-known procedure of histogram equalization, *c.f.* [7]. Histogram equalization is parameter-free and increases the contrast in an image by finding a tone mapping that evens out the intensity histogram of the input image as much as possible. However, for many images histogram equalization gives a too extreme mapping, which for example saturates the brightest intensities so that structure information here is lost. We therefore apply contrast limited histogram equalization as presented by Pizer *et al.* in [18], but without the tiling that applies different mappings to different areas (tiles) in the image. In the contrast limited histogram equalization, a parameter, the clip-limit β , sets a limit on the derivative of the slope of the mapping function. If the mapping function, found by histogram equalization, exceeds this limit the increase in the critical areas is spread equally over the mapping function.

After the noise reduction and tone mapping steps there may still be a need for further contrast enhancement since the adaptive smoothing has a tendency to squeeze the intensities into the central part of the dynamic range, so that the peripheral intensity values are scarcely used. This effect can be alleviated by setting, say, the 0.1 percent of the pixels with the lowest intensity values to zero, which might amount to pixels with intensities values up to between 50 and 75 of 256, and then stretching the colormap uniformly. Applying the same procedure for the brightest part of the intensity range is usually not a good idea, since there are often small bright details, such as a single lamp, that will have a clear change in appearance from the input to the output after such an operation.

5. GPU implementation

The filtering using the summation kernels $k(x_0, x)$ is an inherently parallelizable task for which the graphics processing units of modern graphics cards are very well suited. We have implemented the whole adaptive enhancement methodology as a combined CPU/GPU algorithm.

All image pre- and postprocessing is performed on the CPU. This includes image input/output and the tone mapping step. The histogram equalization, which implements the tone mapping, requires summation over all pixels. This computation is not easily adapted to a GPU, as the summation would have to be done in multiple passes. However, as these steps constitute a small amount of the execution time, a simpler CPU implementation is adequate here.

The most expensive parts are the calculation of the structure tensor, including the gradient calculation, the element-wise smoothing, and the actual filtering, or summation,

which we perform entirely on the GPU. To calculate the gradients we upload n frames of the input sequence to the graphics card as floating-point 2D textures, and perform spatial gradient computations on each of these. Next we take temporal differences of each successive frame and use the resulting gradient to compute the structure tensor for each pixel in each frame. Using the separability of the Gaussian kernel, we then compute the isotropically smoothed tensor for each frame. Typically these filtering steps are performed using filter sizes of up to $7 \times 7 \times 7$ pixels. For the normalization of the filter, the alpha channel is used to store the sum of the filter weights.

The smoothing kernels are then computed in a fragment program on the GPU. This involves finding the eigenvalues and eigenvectors of the structure tensor, which can be done efficiently in a single pass on modern GPUs. The kernel coefficients are temporarily stored as textures and the final spatiotemporal summation is performed, similarly to the isotropic pre-pass, in multiple 2D passes. As the filter kernels are unique for each pixel, the filter weights are re-computed on the fly during the filtering process.

In summary, by exploiting the massively parallel architecture of modern GPUs, we obtain interactive frame rates on a single nVidia GeForce 8800-series graphics card. For some sample RGB sequences, we obtain the following computational times (in ms):

Resolution	CPU computations	Tensor computation	Spatiotemporal smoothing
360×288	14	37	95
1024×768	150	265	702

These timings assume an isotropic pre-smoothing with kernels of size 7^3 pixels, and adaptive spatiotemporal smoothing with filter kernels as large as 13^3 pixels.

6. Results

We have tested our complete enhancement method on a variety of grayscale and color image sequences, differing in resolution and light level, obtained by both consumer grade and machine vision cameras. The tested input sequences include movies obtained by stationary cameras as well as by moving cameras. The method overall gives strikingly clean and bright output, considering the bad quality of the input data.

In Figure 1 one frame of a color sequence obtained by a Sony DCR-HC23 hand-held consumer camera is shown. The upper left image shows the original dark sequence. The resolution of this clip is 360×288 pixels. On the upper right hand side, the tone mapped image is shown, utilizing the contrast limited histogram equalization described in Section 4. The image shown in the lower left corner is taken from the sequence after our extended adaptive smoothing approach in Section 3 has been applied. The smoothing de-



Figure 1. Upper left: One frame of the original sequence. Upper right: Tone mapped version. Lower right: After noise reduction. Lower left: After intensity stretching.



Figure 2. Upper left: One frame the of original sequence. Upper right: Tone mapped version. Lower right: After noise reduction. Lower left: After intensity stretching.

creases the contrast somewhat, but this degradation is automatically alleviated by the stretching procedure discussed at the end of Section 4. The result after this step is shown in the panel in the lower right corner of Figure 1.

Figure 2 shows the algorithm at work on a 351×201 gray level sequence taken by a moving camera. The camera used is a Sony XC-50 machine vision camera. As the camera moved, a person walked from left to right in the scene. The algorithm has no problem in smoothing the intensity in the best possible way for these different motions, and the walking person is well preserved. The method also preserves the sharp edges on the buildings while there is large amounts of smoothing in the homogeneous areas.

In Figure 3 we have applied the algorithm to another sequence obtained by the Sony XC-50 camera. The resolution here is 151×181 pixels. In this figure, one should espe-

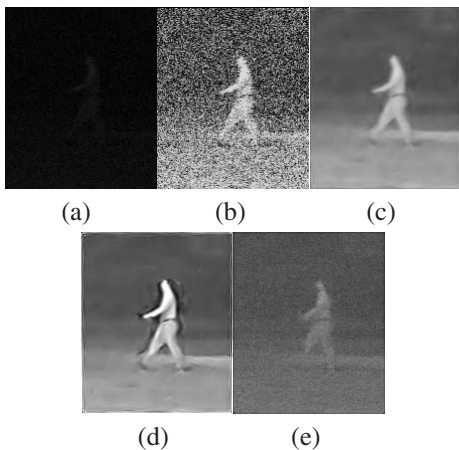


Figure 3. (a) One frame of the original sequence. (b) Tone mapped version. (c) After noise reduction. (d) After noise reduction with sharpening. (e) Noise reduced and tone mapped result using Bennett’s method.

pecially notice the result in panel (d) of applying the sharpening approach as described in Section 3.2, compared to the standard version in panel (c).

6.1. Comparison to other methods

We have compared our method mainly to three other related methods.

Firstly, we implemented the low-light level enhancement method presented by Lee *et al.* [14]. In this method, Poisson noise is removed in stationary areas by calculation of the median in a small spatial neighborhood. So-called false color noise is removed by either choosing the intensity value of the current pixel value or the value of the preceding corresponding pixel, depending on a simple measure of the variance in a neighborhood of the pixel. We have implemented this method and tested it on the same sequences as presented in the last section. However, for the very low light levels that we are dealing with here, and the high noise levels that this implies, the output of the method of Lee *et al.* were visually very visually disappointing, since for instance, the median calculations in a small neighborhood leave heavy flickering in the output.

For a related noise reduction approach, we have implemented the adaptive weighted averaging (AWA) filter of Özkan *et al.* [29], where a weighted mean is calculated within a spatio-temporal neighborhood along the motion trajectories in the sequence. The motion trajectories are calculated using optical flow estimation. For this task, we used the optical flow method suggested by Bruhn *et al.* [4]. However, the heavy noise in our test sequences results in poor estimations of the optical flow and this leads to excessive smoothing in the output of the AWA algorithm. Moreover, much of the noise is left unaffected at the contours of mov-

ing objects in the scene.

The competing method which gave the best output on our dark noisy test sequences was the virtual exposures method as proposed by Bennett and McMillan [3], which was described briefly in Section 2. An output image from this smoothing and tone mapping algorithm is shown in panel (e) in Figure 3. Since this sequence is obtained by a stationary camera, the virtual exposure method mostly here results in temporal filtering which removes a large part of the heavy noise in the input sequence. However, the output isn’t as spatially smooth as the output of the method presented in this paper, as shown in panel (c) and (d). Also, the contrast enhancement achieved by the tone mapping is very disappointing in this case. We suspect that this is due to the fact that Bennett’s tone mapping approach uses separate processing of the low and high frequency parts of the input data which isn’t optimal for the extreme levels of noise that we are dealing with here.

7. Conclusions

In this paper we have presented a methodology for adaptive enhancement and noise reduction for very dark image sequences with very low dynamic range. The approach is very general and adapts to the spatiotemporal intensity structure in order to prevent motion blur and smoothing across important structural edges. The method also includes a sharpening feature which prevents the most important object contours from being over-smoothed. Most parameters can be set generally for a very large group of input sequences. These parameters include: the clip-limit in the contrast-limited histogram equalization, the maximum and minimum widths of the filtering kernels and the width of the isotropic smoothing of the structure tensor and in the gradient calculations. However, the scaling parameter for the width function has to be adjusted to the noise level in the current sequence. The best approach when applying the method to color images has been discussed, which includes demosaicing from the Bayer pattern in raw input color data simultaneously to the noise reduction. We have implemented the method using a GPU and achieved interactive performance. For very noisy video input data, which is the result of filming in very low light levels, the method presented here outperforms (in terms of output quality) all competing methods that we have come across.

References

- [1] F. Alter, Y. Matsushita, and X. Tang. An intensity similarity measure in low-light conditions. In *ECCV ’06*, pages 267–280, 2006. 2
- [2] D. Barash and D. Comaniciu. A common framework for nonlinear diffusion, adaptive smoothing, bilateral

- filtering and mean shift. *Image and Vision Computing*, 22:73–81, 2004. 2
- [3] E. Bennett and L. McMillan. Video enhancement using per-pixel virtual exposures. In *Proc. SIGGRAPH*, pages 845–852, Los Angeles, CA, 2005. 3, 5, 7
- [4] A. Bruhn, J. Wieckert, and C. Schnörr. Lucas/kanade meets horn/schunk: Combining local and global optic flow methods. *Int. J. Computer Vision*, 61(3):211–231, 2005. 7
- [5] F. Drago, K. Myszkowski, T. Annen, and N. Chiba. Adaptive logarithmic mapping for displaying high contrast scenes. In *Proc. EUROGRAPHICS*, volume 22, pages 419–426, 2003. 5
- [6] R. Dugad and N. Ahuja. Video denoising by combining kalman and wiener estimates. In *Proc. Int. Conf. Image Proc.*, volume 4, pages 152–156, Kobe, Japan, October 1999. 2
- [7] R. Gonzalez and R. Woods. *Digital Image Processing*. Addison-Wesley, 1992. 4, 5
- [8] H. Haussecker and H. Spies. *Handbook of Computer Vision and Applications*, volume 2, chapter Motion, pages 125–151. Academic Press, 1999. 4
- [9] K. Hirakawa and T. Parks. Joint demosaicing and denoising. *IEEE Trans. Image Processing*, 15(8):2146–2157, 2006. 4
- [10] B. Jähne. *Spatio-temporal image processing*. Springer, 1993. 3
- [11] D. Kalivas and A. Sawchuk. Motion compensated enhancement of noisy image sequences. In *Proc. IEEE Int. Conf. Acoust. Speech. Signal. Proc.*, pages 2121–2124, Albuquerque, NM, April 1990. 2
- [12] S. Kang, M. Uytendaele, S. Winder, and R. Szeliski. High dynamic range video. In *SIGGRAPH '03: ACM SIGGRAPH 2003 Papers*, pages 319–325, 2003. 2
- [13] S. Lee and M. Kang. Spatio-temporal video filtering algorithm based on 3-d anisotropic diffusion equation. In *Proc. Int. Conf. Image Processing*, volume 2, pages 447–450, 1998. 2
- [14] S.-W. Lee, V. Maik, J. Jang, J. Shin, and J. Paik. Noise-adaptive spatio-temporal filter for real-time noise removal in low light level images. *IEEE Tran. Consumer Electronics*, 51(2):648–653, 2005. 3, 7
- [15] H. Malm and E. Warrant. Motion dependent spatiotemporal smoothing for noise reduction in very dim light image sequences. In *Proc. Int. Conf. Pattern Recognition*, pages 954–959, Hong Kong, 2006. 3
- [16] K. Miyata and A. Taguchi. Spatio-temporal separable data-dependent weighted average filtering for restoration of the image sequences. In *Proc. IEEE Int. Conf. Acoust. Speech. Signal. Proc.*, volume 4, pages 3696–3699, Albuquerque, NM, May 2002. 2
- [17] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. on Patt. Analysis and Machine Intelligence*, 12(7):629–639, 1990. 2
- [18] S. Pizer, E. Amburn, J. Austin, R. Cromartie, A. Geselowitz, T. Geer, B. teer Haar Romeny, J. Zimmerman, and K. Zuiderveld. Adaptive histogram equalization and its variations. In *Computer Vision, Graphics and Image Processing*, volume 39, pages 355–368, 1987. 5
- [19] Y. Reibel, M. Jung, M. Bouhifd, B. Cunin, and C. Draman. Ccd or cmos camera noise characteristics. In *Proc. Euro. Physical J. of Applied Physics*, pages 75–80, 2003. 2
- [20] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec. *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. Morgan Kaufmann, 2005. 2
- [21] M. Sezan, M. Özkan, and S. Fogel. Temporally adaptive filtering of noisy image sequences. In *Proc. IEEE Int. Conf. Acoust. Speech. Signal. Proc.*, volume 4, pages 2429–2432, Toronto, Canada, May 1991. 2
- [22] C. Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking. In *Proc. IEEE Conf. Computer Vision Pattern Recog.*, volume 2, pages 246–252, Fort Collins, CO, June 1999. 3
- [23] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proc. 6th Int. Conf. Computer Vision*, pages 839–846, 1998. 2
- [24] D. Uttenweiler, C. Weber, B. Jähne, R. Fink, and H. Scharf. Spatiotemporal anisotropic diffusion filtering to improve signal-to-noise ratios and object restoration in fluorescence microscopic image sequences. *J. Biomedical Optics*, 8(1):40–47, 2003. 2
- [25] E. Warrant. Seeing better at night: life style, eye design and the optimum strategy of spatial and temporal summation. *Vision Research*, 39:1611–1630, 1999. 1, 3
- [26] J. Weickert. *Anisotropic Diffusion in Image Processing*. Teubner-Verlag, Stuttgart, 1998. 2
- [27] J. Weickert. Coherence-enhancing diffusion filtering. *Int. J. Computer Vision*, 31(2-3):111–127, 1999. 2
- [28] G. Yang, P. Burger, D. Firmin, and S. Underwood. Structure adaptive anisotropic image filtering. *Image and Vision Computing*, 14:135–145, 1996. 2, 3, 4
- [29] M. Özkan, M. Sezan, and A. Tekalp. Adaptive motion-compensated filtering of noisy image sequences. *IEEE Tran. Circuits Sys. Video Tech.*, 3(4):277–290, August 1993. 2, 7