

CONSTRAIN PROPAGATION FOR GHOST REMOVAL IN HIGH DYNAMIC RANGE IMAGES

Matteo Pedone and Janne Heikkilä

Machine Vision Group, Department of Electrical and Information Engineering, University of Oulu, Finland

Keywords: Motion detection, density estimation, image fusion, energy minimization.

Abstract: Creating high dynamic range images of non-static scenes is a challenging task. Carefully preventing strong camera shakes during shooting and performing image-registration before combining the exposures cannot ensure that the resulting HDR image is consistent. This is eventually due to the presence of moving objects in the scene that causes the so called ghosting artifacts. Currently there is no robust solution that produces satisfactory results in any circumstance. Our method consists of two main steps. First, the probability of belonging to the static part of the scene is estimated for each pixel of the N exposures, yielding N weight images. In the second phase, we segment the areas of the weight-images with lower and higher probability values, and smoothly propagate their influence until a significant change in luminosity is detected or a pixel with a corresponding high probability of belonging to the background is approached. This represents an attempt to spread the influence of lower weights to the surrounding pixels of the same object. Results produced with our technique show a significant reduction or total removal of ghosting artifacts.

1 INTRODUCTION

High dynamic range images are commonly created by combining a set of images acquired with varying exposure time (Debevec and Malik, 1997). Ideally, the information contained in the pixels of one exposure must be combined with the corresponding pixels of the other exposures representing the same features of the scene. This is usually accomplished by attempting to avoid camera shakes during shooting and by performing an extra image registration step before merging the exposures. However, although this may maximize the overlap between the exposures, it is not possible to guarantee that the same spatial locations in the low dynamic range images represent the same details of the scene; this is due to the non-static nature of most scenes encountered in practical applications. This situation eventually occurs in the regions of the image space where movement was present or where the image registration algorithm failed to properly align the shots. Combining an exposure set suffering from such inconsistencies will surely cause ghosting artifacts to be visible in the resulting HDR image, in

the form of semi-transparent features. Unless ghosting is involved only in a reasonably small area of the image, it usually produces unacceptable results; this restricts the use of high dynamic range imaging only to static scenes.

2 PREVIOUS WORK

Some authors treat ghosting as an analogous problem to image-registration; a possible solution would then be tracking object movements across the exposures and use this information to warp the pixels of the images accordingly, in order to produce an accurate alignment (Bogoni, 2000). These techniques rely on motion estimation by optical-flow, and it is not possible to ensure they work properly with any kind of movement; for example they inevitably fail when details in some exposures do not have any correspondence in others, due to occlusion.

Ward in (Reinhard et al., 2005) proposes to detect and segment movement regions and replace each segment with information extracted from a single ex-

posure. Movement is detected using a measure based on a weighted variance of pixel values across the exposures, for each location; regions are then selected by marking only the locations where the variance exceeds a threshold value. The main drawback of this method is that ghost regions are replaced by pixels from a single reference image; if a segmented ghost occurs in an area where the dynamic range is high, any exposure will contain underexposed or saturated pixels in that area (or eventually both). Moreover the weighted variance, does not always work properly as a measure for detecting movement; in particular it usually fails when the pixel values of the moving object are similar to those of the background. This motivated Jacobs et al. in (Jacobs et al., 2005) to introduce an auxiliary measure based on entropy that is insensitive to the amount of contrast in the pixel data and efficiently allows to detect movement of features that are similar to the background; nonetheless detected areas are again replaced with pixels of a single exposure.

Grosch in (Grosch, 2006) describes a similar approach based on the idea that the functions that map the intensity of any pixel of one exposure to the corresponding intensity in another exposure can be retrieved, if the camera response function and the exposure times are known. He then uses such functions to predict pixel values across the exposures and create an error map. The error map is thresholded in order to mark invalid locations, that are replaced with pixels from one or possibly more exposures. The algorithm however cannot remove ghosts whose colors are similar to the ones of the background, and the dynamic range in movement areas remains limited.

A different approach is described in (Khan et al., 2006) and it consists in assigning to each pixel of every exposure its probability of belonging to the background. The weights used for pixel values when combining the exposures, are then chosen to reduce the effect of underexposed or saturated pixels as well as to reduce the influence of pixels with a low likelihood of belonging to the background. The procedure used in order to compute these probabilities can be eventually reiterated to get progressively more accurate results. The method, although computationally expensive compared to the others, works fairly well as long as the exposures prevalently capture the background, and it is able to reduce ghosting artifacts in regions where the dynamic range is high. However the algorithm is based on non-parametric kernel density estimation whose efficiency is dependent on a good choice of the smoothing parameters. In addition a density estimation approach is applied to data sets that are not suitable, since analogous elements even-

tually correspond to different values due to variations in overall brightness across the exposures.

3 DETECTING BACKGROUND PIXELS

We treat ghosting as an analogue problem of detecting pixels belonging to the static part of the scene (the background). First, for each location in image space, a statistical model of the background is created, based on information related to the pixels of a neighborhood; successively, the chance of belonging to this class is evaluated for the pixels of interest. A low probability is likely to be caused by the local presence of ghosts. Our approach shares similarities with the iterative scheme suggested in (Khan et al., 2006), however we bring important improvements that make it more suitable for high dynamic range applications and lead to a superior removal of ghosts, as shown in our results.

A three-element vector of the form $(c_1, c_2, c_3) \in \mathfrak{R}^3$ is associated to an arbitrary pixel; its elements correspond to the color information of that pixel. Let R be the total number of exposures; for each spatial location (i, j) the set $N_{i,j}$ of neighboring pixels is used as an approximate representation of the background and it is defined as $N_{i,j} = \{f(x, y, r) \mid r \in [1..R], i-1 \leq x \leq i+1, j-1 \leq y \leq j+1, x \neq i, y \neq j\}$, where $f(x, y, r) \in \mathfrak{R}^3$ denotes the three-components pixel color values at coordinates (x, y) in the r -th exposure. If we think of an exposures sequence as a $M \times N \times R$ image, $N_{i,j}$ can be seen as the set of all the pixels in a rectangular $3 \times 3 \times R$ region in 3d-space excluding all the central pixels. The probability p for a pixel at coordinates (i, j, r) with $r = 1..R$ of belonging to the class $N_{i,j}$ is estimated by a non-parametric kernel density estimation approach. The probabilities $\hat{p}(\bar{\mathbf{x}})$ where $\bar{\mathbf{x}} \in \mathfrak{R}^3$ denotes a generic pixel vector, are then given by

$$\hat{p}(\bar{\mathbf{x}}) = \|\mathbf{H}\|^{-1} \frac{\sum_{i=1}^N w(\bar{\mathbf{X}}_i) K(\mathbf{H}^{-1}(\bar{\mathbf{x}} - \bar{\mathbf{X}}_i))}{\sum_{i=1}^N w(\bar{\mathbf{X}}_i)}. \quad (1)$$

N is the total number of samples, the vectors $\bar{\mathbf{X}}_i$ are the sample pixels that constitute the model of the background, w is a weight function that limits the influence of those pixels that are under-exposed or saturated, and have little chance to be part of the background; the latter is achieved by reusing the values $\hat{p}(\bar{\mathbf{x}})$ from the previous iteration of the algorithm as described in (Khan et al., 2006). K is a d -variate kernel function, and \mathbf{H} is a symmetric positive definite $d \times d$ bandwidth matrix.

Since the particular shape of the kernel function generally does not have a critical impact in performance terms (Silverman, 1986), a standard multivariate Gaussian kernel with identity covariance matrix is often a reasonable choice for K , that is

$$K(\bar{\mathbf{x}}) = \frac{1}{(2\pi)^{\frac{d}{2}}} \exp\left(-\frac{1}{2}\bar{\mathbf{x}}^T\bar{\mathbf{x}}\right) \quad (2)$$

On the other hand it is well known that the use of density estimation approaches can bring useful results only when the *bandwidth values* (or alternatively, *smoothing parameters*) are properly set. Non-suitable bandwidth values are the cause of *undersmoothing* or *oversmoothing* (Silverman, 1986), and for this reason their importance is to be considered vital. Moreover, the “goodness” of the smoothing parameters strongly depends on the nature of the data set in question that could significantly change between different locations in image space. In addition, it must be mentioned that density estimation approaches work correctly as far as same observations correspond to equal vector values in the data set of samples, and this is not the case for high dynamic range applications, due to overall intensity variations across the exposures.

3.1 Pre-Transforming Pixel Values

Density estimation is a reasonable approach as long as analogue observations of a quantity actually correspond to equal sampled values. However the variation in pixel intensity between any pair of exposures can be described by an intensity-mapping-function (IMF) $\tau_{i,j}$ that maps the brightness of an arbitrary pixel in the i -th exposure into the corresponding brightness in the j -th exposure. Often, only the subset of IMFs τ_i between adjacent exposures is considered, where τ_i stands for $\tau_{i,i+1}$. It has been proved that IMFs are theoretically always monotonic increasing (Grossberg and Nayar, 2003). Let’s have a sequence of R exposures and let c be a pixel value; the functions $\Upsilon_c(r) = \tau_{r-1}(\dots(\tau_0(c)))$, where $r \in [1, R]$ and $\tau_0(c) = c$, describe how the intensity of corresponding pixels varies in function of the exposure number. A function Υ_c , given by the composition of monotonic increasing functions, is monotonic increasing too; from this it follows it has an inverse. However IMFs in practice do not describe the accurate correspondence between pixel intensities due to factors such as clamping, quantization errors and noise, that usually have a significant influence in darker and brighter areas of an image. This suggests the introduction of a valid pixel range, typically $V = [20, 240]$, where the mentioned factors can be considered negligible to some extent. Considering only pixel values $c \in V$, and an arbitrary

exposure s , the function

$$\tilde{\Upsilon}_c(r; s) = \begin{cases} \tau_{r-1}(\dots(\tau_s(c))) & r > s \\ c & r = s \end{cases}$$

can be considered monotonic increasing in the interval $s \leq r \leq l$, where $l = \max\{r \mid \tilde{\Upsilon}_c(r; s) \in V\}$. This intuitively suggests that intensities of corresponding well exposed pixels can and should be transformed into the same value.

We only experimented with a naive approach based on histogram matching. We choose the “best” reference exposure of the sequence and perform a “relaxed” intensity alignment to it for all the other exposures. We typically choose the exposure whose integral of the luminance channel histogram in the intervals $[0, 20]$ and $[240, 255]$ is lowest, preferring the one that uses more bins. Intensity alignment is performed by simple histogram matching, although for this application a more robust radiometric alignment approach like the one described in (Kim and Pollefeys, 2004) could be eventually used. We reduce the loss of information due to intensity alignment, by introducing a simple measure of its “harmfulness”. Let’s have two images A, B and denote by $A \rightarrow B$ the resulting image of an intensity-alignment of A to B (A matches B). We attempt to quantify the loss of information due to the intensity-alignment operation by

$$loss_{A,B} = \frac{|\{hist_L((A \rightarrow B) \rightarrow A) > 0\}|}{|\{hist_L(A) > 0\}|}$$

where $hist_L(X)$ is the luminance channel histogram of an image X , and $|\{hist_L(X) > 0\}|$ returns the total number of used bins. This is justified by the fact that transforming an image A into $A \rightarrow B$ would most likely cause a certain amount of pixels to be clamped to the valid color range, and transforming $A \rightarrow B$ back to its original intensity would yield a resulting image $(A \rightarrow B) \rightarrow A$ suffering from posterization artifacts. The quantity $loss_{A,B}$ is in the range $(0, 1]$ and it is used to define a linear interpolation factor t between the original image and the intensity-matched version of it, in the following way,

$$t = \begin{cases} k & k \leq 1 \\ 1 & k > 1 \end{cases}$$

where

$$k = \frac{loss_{A,B}}{1-tolerance} \quad 0 \leq tolerance < 1$$

and the factor *tolerance* determines when to start to linearly interpolate. We set by default *tolerance* = 0.3. Density estimation is finally performed on the set of intensity-aligned images (with opportune “relaxation”), while the output weights of the algorithm are of course used with the original set of exposures.

3.2 Estimating Bandwidth Matrices

In (Raykar and Duraiswami, 2006) Vikas et al. describe methods to quickly compute the optimal bandwidth value for univariate data, and in (Zhang et al., 2004) an algorithm to estimate full bandwidth matrices in the multivariate case is proposed. A natural approach would then be applying such algorithms to the pixel data contained in $N_{i,j}$, and repeating this step at each location (i, j) ; nevertheless the amount of calculations required would become unacceptable. Using a decorrelated color space like $L\alpha\beta$ or CIELAB makes the choice of a *product kernel* (\mathbf{H} diagonal) reasonable. This led us to consider the possibility of estimating the bandwidth parameters using a straightforward and fast method. We used a slightly modified version of Scott’s rule (Scott, 1992), that in its original form is

$$\tilde{h}_i \approx n^{-\frac{1}{d+4}} \sigma_i$$

where d is the dimensionality of the n samples, and σ_i is their standard deviation in the i -th dimension. Concerning the terms σ_i , we prefer to compute *weighted standard deviations* of the following form,

$$\tilde{\sigma} = \sqrt{\frac{\sum_{i=1}^n w(x_i) \cdot (x_i - \bar{x})^2}{\sum_{i=1}^n w(x_i)}}$$

where $w(x_i)$ are the sample weights, \bar{x} is the weighted mean of the samples x_i , and n is the total number of samples. We use the pixel weights obtained from the last iteration of the algorithm to reduce the influence of the pixels that are incorrectly exposed and have little chance of belonging to the background. We finally propose the following bandwidth matrix for an arbitrary neighborhood set $N_{i,j}$:

$$\mathbf{H}_{i,j} = |N_{i,j}|^{-\frac{1}{2}} \text{diag}(\tilde{\sigma}_L(N_{i,j}), \tilde{\sigma}_\alpha(N_{i,j}), \tilde{\sigma}_\beta(N_{i,j})) \quad (3)$$

where we respectively denoted by $\tilde{\sigma}_L(N_{i,j})$, $\tilde{\sigma}_\alpha(N_{i,j})$ and $\tilde{\sigma}_\beta(N_{i,j})$ the weighted standard deviations of the luminance and color data of the pixels of the neighborhood $N_{i,j}$.

In addition, it is possible to obtain satisfactory results and save computation time by precalculating the set of matrices $\mathbf{H}_{i,j}$ in the beginning and reusing it at every iteration, or alternatively by applying (3) to the entire set of pixels of all the exposures in order to obtain only one global suitable bandwidth matrix.

3.3 Weights Propagation

It often happens that although some portions of a moving object are correctly detected and assigned low probabilities of belonging to the static part of the

scene, other parts are still given higher weights. This can be due to several reasons like similarity of object and background colors, or more generally to limitations in the density estimation approach. Intuitively, if a portion of a moving object has low chances of belonging to the background, this should be eventually valid also for all the other portions of the same object. We describe a method that attempts to propagate the influence of the lower probabilities to the surrounding areas of the image representing the same feature. After each iteration of the procedure described in the previous sections, a new $M \times N \times R$ matrix of normalized weights is available. Before merging the exposures, we compute two threshold values to segment the areas with relatively low and high likelihood of being part of the static part of the scene; the values we use are respectively the 10-th percentile and the 60-th percentile of all the pixel weights. For each exposure, two morphological operation are applied to the binary images relative to the lower and higher weights: respectively *close* and *open* using a disc shaped structuring element of 7 pixels radius. The former image is multiplied by a low value (we use 0.01), and its nonzero pixels replace the corresponding ones in the latter image. This procedure yields R constrain-images that can be used in conjunction with the original exposures to perform an image-driven propagation; the approach used is the one described in (Lischinski et al., 2006) that minimizes the following quadratic functional:

$$f = \arg \min_f \left\{ \sum_{\mathbf{x}} w(\mathbf{x}) (f(\mathbf{x}) - g(\mathbf{x}))^2 + \lambda \sum_{\mathbf{x}} \left(\frac{|f'_x(\mathbf{x})|^2}{|L'_x(\mathbf{x})|^\alpha + \varepsilon} + \frac{|f'_y(\mathbf{x})|^2}{|L'_y(\mathbf{x})|^\alpha + \varepsilon} \right) \right\}$$

Intuitively the first term aims to keep the resulting pixel weights given by f , as close as possible to the values $g(\mathbf{x})$ of the constrain-image, while the second term is necessary to keep the gradient of the objective function small, allowing however large changes when the gradient of the log-luminance channel L of the underlying image is significant. The weights $w(\mathbf{x})$ and the parameters λ , α , and ε can be used to control the type of propagation; in our experiments we used the default values $\lambda = 0.2$, $\alpha = 1$, $\varepsilon = 0.0001$, while $w(\mathbf{x})$ is equal to 1 in correspondence of the nonzero values of the constrain image, and 0 otherwise. The propagated weights are then multiplied element by element by the original weight matrix, and used either to combine the exposures, or to perform an extra iteration of the algorithm (Figure 1). Lischinski et al. in their paper describe a fast method that is able to minimize the functional f in a fraction of second for a 640×400 image. In order to save memory and com-

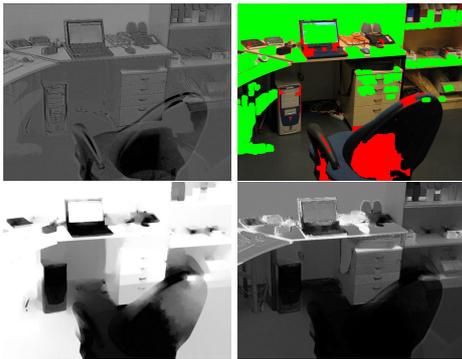


Figure 1: The main steps of the propagation process: weights obtained from density estimation; low and high weights segmented; propagated weights; final weights.

putation time, in case of big images, we perform the propagation step on downsampled versions of the exposures, and resize the results back to their original dimensions.

4 RESULTS AND CONCLUSION

We compared our approach to the one described in (Khan et al., 2006). Reinhard's photographic operator was used to tonemap the generated HDR images (Reinhard et al., 2002). Our approach does not require any setting to be adjusted by the user. In all the experiments shown we used only one global bandwidth matrix that is reused at every iteration: the more general approach described resulted in a significant increase of computation time with little benefits. For Khan's method, we used a default identity bandwidth matrix, and $3 \times 3 \times R$ neighborhoods. We included in Figure 2 some of the exposures used for generating the final HDR images. Figure 3 shows the results of the experiments. In the first scene, ghosting is localized and occurs in regions that have high dynamic range; artifacts are completely removed only with our algorithm. In the second scene, the situation is similar but less exposures were available. Density estimation alone could not distinguish properly the background, while the weight propagation helped to improve the results. Finally we considered a handheld set of exposures intentionally left unaligned, and where chaotic movement is present; this sequence does not hold the assumption that the background is prevalently captured and suffers from critical occlusion and parallax problems. In spite of this, our method proved a remarkable robustness against feature misalignments. In all the cases that have been considered, our approach showed a significant improvement in reducing ghosting arti-

facts, and when the previously mentioned assumption holds, ghosts can be completely eliminated even with a single iteration.

REFERENCES

- Bogoni, L. (2000). Extending dynamic range of monochrome and color images through fusion. In *International Conference on Pattern Recognition, 2000*, vol. 3, pp. 7-12.
- Debevec, P. and Malik, J. (1997). Recovering high dynamic range radiance maps from photograph. In *SIGGRAPH 97, August 1997*.
- Grosch, T. (2006). Fast and robust high dynamic range image generation with camera and object movement. In *Vision, Modeling and Visualization (VMV), RWTH Aachen, 22.11 - 24.11.2006*.
- Grossberg, M. D. and Nayar, S. K. (2003). Determining the camera response from images: What is knowable? In *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.25, No.11, pp.1455-1467, Nov, 2003*.
- Jacobs, K., Ward, G., and Loscos, C. (2005). Automatic hdri generation of dynamic environments. In *ACM SIGGRAPH 2005 Sketches*.
- Khan, E. A., Akyuz, A. O., and Reinhard, E. (2006). Ghost removal in high dynamic range images. In *IEEE International Conference on Image Processing, Atlanta, USA, August 2006*.
- Kim, S. J. and Pollefeys, M. (2004). Radiometric alignment of image sequences. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04) - Volume 1, pp. 645-651*.
- Lischinski, D., Farbman, Z., Uyttendaele, M., and Szeliski, R. (2006). Interactive local adjustment of tonal values. In *ACM Transactions on Graphics, ACM SIGGRAPH 2006 Papers SIGGRAPH '06, Volume 25 Issue 3*. ACM Press.
- Raykar, V. C. and Duraiswami, R. (2006). Very fast optimal bandwidth selection for univariate kernel density estimation. In *CS-TR-4774, Department of Computer Science, University of Maryland, Collegepark*.
- Reinhard, E., Stark, M., Shirley, P., and Ferwerda, J. (2002). Photographic tone reproduction for digital images. In *ACM Transactions on Graphics, 21(3), pp 267-276, Proceedings of SIGGRAPH 2002*.
- Reinhard, E., Ward, G., Pattanaik, S., and Debevec, P. (2005). *High Dynamic Range Imaging: Acquisition, Display and Image-Based-Lighting*. Morgan Kaufmann.
- Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.
- Zhang, X., King, M. L., and Hyndman, R. J. (2004). Bandwidth selection for multivariate kernel density estimation using mcmc. In *Monash Econometrics and Business Statistics Working Papers 9/04*.



Figure 2: Some of the exposures used in the experiments. From left to right, the original sequences contained respectively seven, five, and six exposures.



Figure 3: Results obtained without any ghost removal applied (left column); with Khan's method (middle column); with our method (right column). Six iterations of Khan's method have been used for the first two sequences, and four iterations for the third one; further iterations did not improve significantly the final image. All the results produced with our method required only a single iteration. In the first sequence, ghosting artifacts of the car on the right are completely removed only with our algorithm. The propagation approach works fairly well also with ghosting artifacts caused by feature misalignments.