

## GRADIENT DOMAIN CONTEXT ENHANCEMENT FOR FIXED CAMERAS\*

ADRIAN ILIE

*Computer Science Department  
University of North Carolina at Chapel Hill  
Sitterson Hall, CB# 3175, Chapel Hill, NC 27599, USA  
adyilie@cs.unc.edu*

RAMESH RASKAR

*Mitsubishi Electric Research Labs  
201 Broadway, 8th Floor, Cambridge, MA 02139, USA  
raskar@merl.com*

JINGYI YU

*Laboratory of Computer Science  
Massachusetts Institute of Technology  
200 Technology Square, Cambridge, MA 02139, USA  
jingyi@graphics.csail.mit.edu*

We propose a class of enhancement techniques suitable for scenes captured by fixed cameras. The basic idea is to increase the information density in a set of low quality images by extracting the context from a higher-quality image captured under different illuminations from the same viewpoint. For example, a night-time surveillance video can be enriched with information available in daytime images. We also propose a new image fusion approach to combine images with sufficiently different appearance into a seamless rendering. Our method ensures the fidelity of important features and robustly incorporates background contexts, while avoiding traditional problems such as aliasing, ghosting and haloing. We show results on indoor as well as outdoor scenes.

*Keywords:* Image enhancement; video enhancement; gradient domain integration.

### 1. Introduction

In recent years, a number of techniques have emerged to extract useful information from multiple images taken from a fixed viewpoint. They include video summarization, generation of intrinsic images, multispectral image fusion and high dynamic range (HDR) compression.

\*This paper is an invited submission of revised material previously published in the proceedings of ACCV 2004.<sup>8</sup> It contains more detailed presentations of our methods, as well as expanded versions of the algorithms.

In this paper we propose a different class of image and video enhancement techniques, which we call *context enhanced rendering* (CER, for simplicity and because there is no common term). The goal of CER is to extract scene information from one image and use it as context for important features from another image of the same scene. HDR imaging and image fusion are special cases of CER. We call the image that provides environmental context the *background image*, and the one that provides desirable features the *foreground image*.

A typical example of CER we consider in this paper is enhancing night-time traffic or surveillance videos using daytime images taken from the same viewpoint. Usually a night-time video is very difficult to understand because it lacks background context due to poor illumination. However, the elements of this background context, such as roads and buildings are critical to understanding the video as shown in Figs. 1 and 2. While a trained traffic controller may easily recognize important features in night-time videos, we think that our method would help nonspecialists achieve the same performance. Moreover, even traffic controllers may benefit

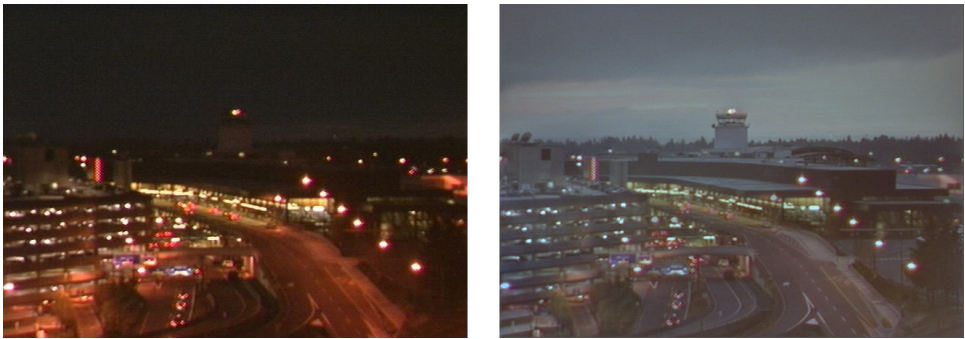


Fig. 1. Enhancing a night-time scene from an airport surveillance camera. A low quality night-time image, and the final output of our algorithm.



Fig. 2. Enhancing a traffic video. A low quality night-time video frame, and the final output of our algorithm.

from our approach: they may confirm their suspicions by switching between normal imagery and our context-enhanced imagery.

### 1.1. Overview

Our approach is a new image fusion approach to combine snapshots of the same scene with significantly different appearance into a seamless rendering. For the rest of the paper, we limit ourselves to this type of CER. The method maintains the fidelity of important features and robustly incorporates background contexts, while avoiding problems encountered in traditional methods, such as aliasing, ghosting and haloing. We compute the importance of a pixel from the local variance in the input snapshots or videos. Afterwards, instead of a convex combination of pixel intensities, we combine the intensity gradients scaled by the importance. By reconstructing the result image through integration of the combined gradients, our method achieves a smooth blend of the inputs, and at the same time preserves their important features.

Just as the compression of HDR images, the goal of CER should be to produce “visually pleasing” results, with very few artifacts and a smooth transition from background to foreground. Our method accomplishes this by using the underlying properties of integration. We show how this can be used for synthetic as well as natural indoor and outdoor scenes.

A common artifact of gradient-based reconstruction is that it introduces observable color shifting. We discuss in detail the causes of these artifacts and show a color assignment scheme that can efficiently reduce them.

### 1.2. Contributions

Our main contribution is the idea of exploiting information available from fixed cameras to create context-rich, seamless results. Our technical contributions include a scheme for asymmetrically fusing two snapshots of the same scene while preserving useful features; and a method for context enhancement of videos in the presence of unreliable frame differencing.

In addition, we modify the reconstruction from gradients method with a padding scheme to overcome common integration artifacts such as aliasing, ghosting and haloing; and we use a color assignment strategy to address color shifting problems.

### 1.3. Related work

Methods to combine information from multiple images into a single result have been explored for various other applications. They range from image editing<sup>11</sup> and tone mapping for compression of variable-exposure high-dynamic range images,<sup>5,13</sup> to art such as “Nude on a step” by Duchamp and “The Empire of Light” by René Magritte.

The authors of Ref. 1 use multiresolution splines for combining images into a seamless image mosaic. The source images are first decomposed into a set of

band-pass filtered component images. Next, the component images in each spatial frequency band are assembled into a corresponding band-pass mosaic using a weighted average within a transition zone which is proportional in size to the wavelengths represented in the band. Finally, these band-pass mosaic images are summed to obtain the desired image mosaic.

In HDR imaging, a set of images taken under different levels of exposure are combined into a single image where details in all of the images are preserved while the overall contrast is reduced. However, in HDR, the pixel intensities increase monotonically. Usually cameras can only capture images one fixed exposure at a time, hence can only capture part of the scene when the radiance range of the scene is large. The goal of HDR is to compensate for context that is missing in one exposure setting by using information from another. Two classes of approaches have been suggested: image space<sup>4</sup> and gradient space<sup>5</sup> methods. A recent approach<sup>9</sup> combines classic HDR techniques with motion estimation and other video methods to obtain HDR video.

While its goals are similar, our problem is quite different from combining HDR images. For example, in combining day-night images we encounter intensity gradient reversals (such as objects that are darker than their surroundings during the day, but brighter than their surroundings during the night). An example of such a reversal is a building that is lit during the night so it becomes brighter than the night-time sky, yet during the day it is darker than the day-time sky. These reversals do not appear in HDR, but need to be dealt with in general CER methods.

Another example of CER is image fusion for multispectral imagery e.g. to merge satellite imagery captured at different wavelengths. Here, the images are relatively similar. Many ideas from multispectral image fusion can mutually benefit CER. Our approach is closest to the one proposed in Ref. 15. They put forward a gradient space method by first forming a unified gradient image and then searching for an optimal image that satisfies the gradient image.

A similar problem to enhancing images with context is removing or reducing undesirable context in images, such as shadows or fog. Authors of Ref. 6 removed shadows in an image by first computing its gradient, then distinguishing shadow edges, setting the gradient values at the shadow edges to zero and finally reintegrating the image. Nayar *et al.* used time-lapsing image sequences to model the effect of fog.<sup>10</sup> By setting appropriate parameters, they are able to efficiently enhance images and reduce undesirable weather artifacts.

Image and video matting<sup>2</sup> uses Bayesian models for robust foreground segmentation. Continuous blending methods are employed to integrate the foreground and the background e.g. using distance fields.

Pèrez *et al.*<sup>11</sup> presented a technique that uses integration of modified gradients from several images to produce one seamless result. However, since their goal is to provide a framework for seamless image editing, they rely heavily on user input to assign the areas from which the gradients are taken. The user designates which

areas should come from which image, which is equivalent to a particular case of our method with simple or no blending and precise manual segmentation.

We believe robust foreground extraction in image space is difficult to achieve in practice, especially in real-time or when dealing with low contrast and noisy snapshots and videos. Therefore we propose a gradient space algorithm that avoids a lot of undesirable artifacts like aliasing, ghosting and haloing that appear when using conventional methods, while being simple enough to warrant a future hardware implementation. Our algorithm consists of two major steps similar to video matting: foreground extraction and background fusion.

We demonstrate our algorithm in different situations and show that our method is robust to poor background segmentation and generates results with very few artifacts. We are inspired by many of the techniques mentioned here and aim to address some of their limitations.

## 2. Basic Technique

To illustrate our new image fusion approach of combining snapshots of the same scene with sufficiently different appearance into a seamless rendering, we focus on enhancing poor-context night-time snapshots or videos with context elements from high-quality daytime snapshots. This section describes our basic fusion technique. We first present the basic algorithm, then our approach to ensure better reconstruction and color assignment.

### 2.1. Basic algorithm

Our method combines information from two snapshots in a meaningful way, by picking high-quality background information from a daytime snapshot and using it to enhance the low-quality but important information from a night-time snapshot. A straightforward approach is to use a linear combination of the input snapshots. Instead, we specify the desired local attributes of the final result and solve the inverse problem of obtaining a global solution that satisfies the desired attributes. This leads to a nonlinear combination, which means pixels with the same intensities map to different intensities in the final result.

Our method for determining the important areas of each snapshot relies on the widely accepted assumptions<sup>3</sup> that the human visual system is not very sensitive to the absolute luminance reaching the retina, but rather responds to local intensity ratio changes. Hence, the local attribute that we use is the local variance. We define an importance function for each input snapshot based on the spatial intensity gradients, which are a measure of the local spatial variance.

We apply two heuristics to decide what information to carry into the desired result: (a) we take the gradients from the night-time snapshot that appear to be locally important and (b) we use gradients from the daytime snapshot to provide context to locally-important areas while maintaining intra-image coherence. Our method does not improve the quality of the pixels themselves, it simply gives

sufficient context to improve human interpretation. Consequently, operations such as contrast enhancement, histogram equalization, mixed Gaussian models for background estimation<sup>16</sup> are orthogonal to our approach and can be easily used alongside to improve the final result.

The regions of high spatial variance across each snapshot are computed by thresholding the intensity gradients,  $G = (G^X, G^Y)$ , for the horizontal and vertical directions using a simple forward difference. We then compute an *importance image* (a weighting function)  $W$ , by processing the gradient magnitudes  $|G_D|$  and  $|G_N|$  of the daytime snapshot  $D$  and the night-time snapshot  $N$ , respectively. The weighted combination of the input gradients gives us the gradient of the desired output. The basic steps are as described in Algorithm 1 and illustrated in Fig. 3.

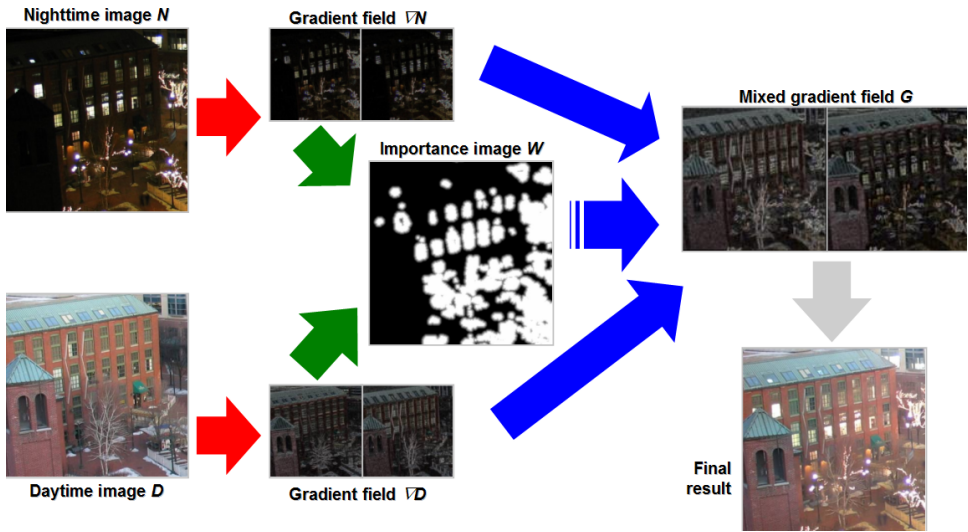


Fig. 3. The basic algorithm.

As described in the following sections, the process of determining importance weights,  $W_{(x,y)}$ , depends on the specific application.

---

**Algorithm 1** Basic algorithm

---

Find gradient field of daytime snapshot  $G_D = \nabla D$

Find gradient field of night-time snapshot  $G_N = \nabla N$

Compute importance image  $W$  from variances  $|G_D|$  and  $|G_N|$

Compute mixed gradient field at each pixel

$$G_{(x,y)} = G_{N(x,y)}W_{(x,y)} + G_{D(x,y)}(1 - W_{(x,y)})$$

Reconstruct result  $I'$  from gradient field  $G$

Normalize pixel intensities in  $I'$  to closely match  $N_{(x,y)}W_{(x,y)} + D_{(x,y)}(1 - W_{(x,y)})$

---



Fig. 4. Overcoming integration artifacts by padding. (Left to right) The original image, the integration result without padding, and the integration result using padding.

## 2.2. Image reconstruction

Image reconstruction from gradients fields is an approximate invertibility problem, and still a very active research area. In 2D, a modified gradient vector field  $G$  may not be integrable. We use one of the direct methods recently proposed<sup>5</sup> to minimize the error  $|\nabla I' - G|$ . The estimate of the desired intensity function  $I'$ , so that  $G = \nabla I'$ , can be obtained by solving the Poisson differential equation  $\nabla^2 I' = \text{div}G$ , involving a Laplace and a divergence operator. We use the full multigrid method<sup>12</sup> to solve the Laplace equation. We pad the input images to square images of size the nearest power of two before applying the integration, and then crop back the result to the original size.

The solver needs to have the boundary conditions specified (at the border of the image). A natural choice is Neumann condition  $\nabla I' \cdot n = 0$  i.e. the derivative in the direction normal to the boundary is zero. This is clearly not true when high gradients are present near the image boundary, resulting in noticeable color bleeding and shifting artifacts. The padding of the image to the nearest power of two that is necessary for multigrid integration also helps alleviate this problem. Figure 4 shows a comparison of integrating the gradient field of an image with and without padding.

The integration of the gradient field involves a scale and shift ambiguity,  $I''_{(x,y)} = c_1 I'_{(x,y)} + c_2$ . To obtain the final image,  $I''$ , we compute the unknowns,  $c_1$  and  $c_2$ , (in the least square sense) using a simple heuristic: the intensity of each pixel of the reconstructed image should be close to the corresponding combination of the pixel intensities of the foreground and background images. Each pixel leads to a linear equation,  $N_{(x,y)}W_{(x,y)} + D_{(x,y)}(1 - W_{(x,y)}) = c_1 I'_{(x,y)} + c_2$ . We do image reconstruction in all three color channels separately and compute the unknowns per channel.

## 3. Enhancement of Dynamic Scenes

For dynamic scenes, our results are based on the observation that if the camera and most of the viewed geometry remain static, the only changes are the illumination

and the moving parts of the scene (e.g. people, devices, vehicles). Thus, the intensity gradients corresponding to the stationary parts in a night-time snapshot can be replaced with better quality gradients from a daytime snapshot.

We use the notion of a static background to provide context for dynamic actions or events happening in the foreground. The static component can be captured at high resolution, under controlled illumination conditions. The dynamic component can be captured in multiple snapshots of lower quality. A good example of applying our technique is enhancing pictures of theme park visitors taken during a ride through a dark environment, when bright flashes cannot be used because they may harm the visitors' eyes. The static background can be inserted from a snapshot captured using brighter illumination, when there are no visitors in the scene. Also, using a higher resolution background image can increase the perceived resolution of the dynamic foreground.

To compute the desired gradient field, the authors of Ref. 11 made the choice of using the local maximum of the input gradients for the weights  $W_{(x,y)}$ , i.e.  $G_{(x,y)} = \max(G_{d(x,y)}, G_{n(x,y)})$ . In this case importance weights are either 0 or 1. A better choice for our application is to give more importance to night-time gradients in the region of the night-time snapshot where gradients or intensities are above a fixed threshold. This is to make sure that no information in the night-time snapshot is lost in the final result.

To provide context to foreground changes in the illumination and geometry of a series of night-time snapshots, we replace low-detail background areas using data from the daytime snapshot. This is where many of the traditional methods using linear combination will fail to create seamless results. Let us consider the case where we want to provide context to a night-time snapshot  $N$  using information from another night-time reference snapshot  $R$  and a daytime snapshot  $D$ . We create

---

**Algorithm 2** Algorithm for enhancing dynamic scenes

---

Compute mask image  $M_{(x,y)} = |N_{(x,y)} - R_{(x,y)}|$  at each pixel

Threshold and smooth mask image  $M$

Find gradient field of daytime snapshot  $G_D = \nabla D$

Find gradient field of night-time snapshot  $G_N = \nabla N$

Compute importance image  $W$  from variances  $|G_D|$  and  $|G_N|$

Compute aggregate weighting function  $W'_{(x,y)} = W_{(x,y)} + M_{(x,y)}$  at each pixel

Normalize the weighting function  $W'$

Compute mixed gradient field at each pixel

$$G_{(x,y)} = G_{N(x,y)} W'_{(x,y)} G_{D(x,y)} (1 - W'_{(x,y)})$$

Reconstruct result  $I'$  from gradient field  $G$

Normalize pixel intensities in  $I'$  to closely match  $N_{(x,y)} W'_{(x,y)} + D_{(x,y)} (1 - W'_{(x,y)})$

---



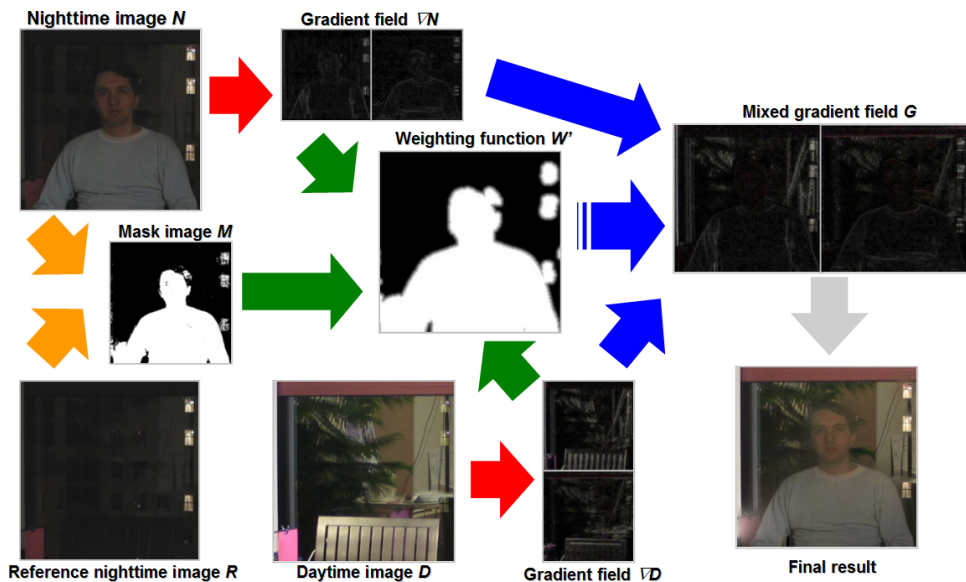


Fig. 5. The algorithm for enhancing dynamic scenes.

a mask image  $M$ , and set  $M_{(x,y)} = |N_{(x,y)} - R_{(x,y)}|$  so that the importance is scaled by the difference between the two night-time snapshots, thus accounting for the areas with changes (temporal variance). Mask  $M$  is thresholded and smoothed, then added with the weighting function  $W$ , which accounts for the areas with high spatial variance. Algorithm 2 (illustrated in Fig. 5) presents the steps in order.

Although we use a very simple segmentation technique (pixel-wise difference in color space between snapshots  $N$  and  $R$ ) to detect important changes at nighttime, our method is robust and does not need to rely on complicated segmentation techniques to obtain reasonable results (Fig. 6 shows the example in Fig. 5 at a higher resolution). This is because we need to detect the difference between  $N$  and  $R$  only where gradients of  $N$  are sufficiently large. In a pair of snapshots, flat regions may have similar color but they naturally differ in regions of high gradient.

We allow for graceful degradation of the result when the underlying computer vision methods fail. More sophisticated segmentation techniques would bring marginal improvements to our results. Additionally, user input can help guide the algorithm by manually modifying the importance image.

#### 4. Enhancement of Videos

We also apply our technique to enhance low quality videos, such as the ones obtained from security and traffic surveillance cameras. In such videos, enhanced context can help answer questions such as: why is a person standing near a part of a building



Fig. 6. Enhancing a dynamic scene. (Top row) A high quality daytime snapshot, a low quality night-time reference, and with a foreground person. (Bottom row) A simple binary mask, the importance image obtained after processing, the final output of our algorithm.

(they are looking at a poster), what is the person's hand hidden by (they are behind a dark object that is not illuminated), what are the reflections in the dark areas (car headlights reflecting from windows of dark buildings), what is a blinking light (traffic light clearly seen at daytime).

The static background, as in the previous section, comes from a single higher-quality daytime snapshot. The dynamic foreground is composed of regions of high variance, both spatial and temporal. As a straightforward extension of the algorithm for dynamic scenes, regions of high temporal variance between two video frames are computed by comparing the intensity gradients of corresponding pixels from the two frames.

Videos present several additional challenges: (a) inter-frame coherence must also be maintained, i.e. the weights in successive frames should change smoothly and (b) a pixel from a night-time frame may be important even if the local variance is small (e.g. the area between the headlights and the tail-lights of a moving car). Our solution is based on the simple observation that in a sequence of video frames, moving objects span approximately the same pixels from head to tail. For example, the front of a moving car covers all the pixels that will be covered by rest of the car in subsequent frames. Using temporal hysteresis, although the body of a car may not show enough intra-frame or inter-frame variance, we maintain the importance

---

**Algorithm 3** Algorithm for enhancing videos

---

Compute spatial gradients for daytime  $G_D = \nabla D$   
 Smooth video using SUSAN  
**for** each video frame  $F_i$  **do**  
   Compute spatial gradients  $G_{Ni} = \nabla F_i$   
   Threshold temporal differences into binary masks  $M_i$   
   Create weights  $W_i$  using  $M_i, |G_D|$  and  $|G_{Ni}|$   
**end for**  
**for** each weight image  $W_i$  **do**  
   Average into  $W'_i$  over  $2c+1$  time steps  
     
$$W'_{k(x,y)} = (\sum_{i=k-c}^{k+c} W_{i(x,y)}) / (2c + 1)$$
  
**end for**  
**for** each video frame  $F_i$  **do**  
   **for** each pixel  $(x,y)$  **do**  
     **if**  $W'_{i(x,y)} > 0$  **then**  
       Compute mixed gradient field as  
         
$$G_{(x,y)} = G_{Ni(x,y)} W'_{i(x,y)} + G_{D(x,y)} (1 - W'_{i(x,y)})$$
  
     **else**  
       Compute mixed gradient field as  $G_{(x,y)} = \max(G_{Ni(x,y)}, G_{D(x,y)})$   
     **end if**  
   **end for**  
   Reconstruct frame  $F'_i$  from gradient field  $G$   
   Normalize pixel intensities in  $F'_i$  to closely match  
     
$$F'_{i(x,y)} W'_{i(x,y)} + D_{(x,y)} (1 - W'_{i(x,y)})$$
  
**end for**

---

weight high in the interval between the head and the tail. The steps are as described in Algorithm 3 and illustrated in Fig. 7.

The importance is based on the spatial and temporal variation as well as the hysteresis computed at a pixel. A binary mask  $M_j$  for each frame  $F_i$  is calculated by thresholding the difference with the previous frame,  $|F_i - F_{i-1}|$ . To maintain temporal coherence, we compute the importance image  $W_j$  by averaging the processed binary masks  $M_k$ , for frames in the interval  $k = i-c..i+c$ . We chose the extent of influence  $c$ , to be five frames in each direction. Thus, weight due to temporal variation  $W_i$  is a mask with values in  $[0,1]$  that vary smoothly in space and time. Then for each pixel of each frame, if  $W_{i(x,y)}$  is nonzero, we use the method of context enhancement of dynamic scenes i.e. blend the gradients of the night-time frame and daytime snapshot scaled by  $W_{i(x,y)}$  and  $(1 - W_{i(x,y)})$ . If  $W_{i(x,y)}$  is zero, we revert to a special case of the method of enhancement for static scenes i.e. choose the gradient with the larger magnitude. Finally, each frame is individually reconstructed

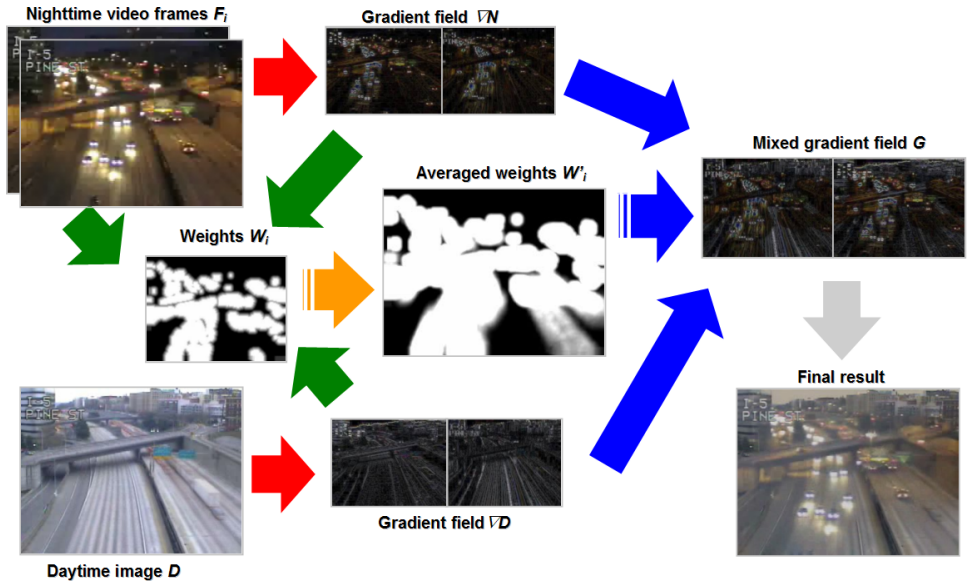


Fig. 7. The algorithm for enhancing videos.

from the mixed gradient field for that frame. Figure 7 shows a visual description of the algorithm, and Fig. 8 shows the same example at a higher resolution.

The noise in the input video is reduced by using a feature-preserving bilateral filtering in three dimensions (space and time). This eliminates false positives when frame-differences are computed. For a practical implementation we repeatedly applied a 3D SUSAN filter<sup>14</sup> ( $3 \times 3 \times 5$  neighborhood,  $\sigma = 15$  and  $t = 20$ ). The daytime snapshot used for filling in the context is obtained by median filtering a daytime video clip (about 15 seconds).

Just as in the case of dynamic scenes, a good quality video segmentation or optical flow technique will marginally improve our results. We intentionally use a very simple technique (pixel-wise difference) to show that the result of our techniques does not need to rely completely on complicated optical flow or image change detection techniques. User input can also easily be incorporated in the process. Since the camera position is static, the user can either designate areas to be filled from the daytime image for all frames, or for each frame separately.

## 5. Discussion

### 5.1. Comparison

We introduced a practical method for improving a low-quality night-time image by combining it with a high-quality daytime scene. This idea appears to be very simple in retrospect. However, despite our search efforts, the idea appears to have been unexplored in image enhancement.



Fig. 8. Enhancing traffic video. (Top row) A high quality daytime and a low quality night-time video frame. (Bottom row) The importance image obtained after processing, and the final output of our algorithm. Notice the road features and background buildings.

A naïve approach to automatically combine a daytime and night-time snapshot would be to use a pixel substitution method based on some importance measure. This works well only when the inputs are almost identical (e.g. two snapshots of the same scene with different focus<sup>7</sup>). Similarly, blending strategies such as  $\max_i(I_i(x,y))$  or  $\text{average}_i(I_i(x,y))$  also create problems. For example, when combining day-night snapshots, one needs to deal with high variance in daytime snapshots and with mostly low contrast and patches of high contrast in night-time snapshots. Taking the average simply overwhelms the subtle details in the night-time snapshot, and presents “ghosting” artifacts around areas that are bright at night-time. Furthermore, juxtaposing or blending pixels usually leads to visible artifacts (e.g. sudden jumps from dark night pixels to bright day pixels) that distract from the subtle information conveyed in the night snapshots. Figure 9 shows a comparison of our method (shown in Fig. 6 with averaging pixel values, as well as blending pixel values using the same importance function we used for blending the gradients).

## 5.2. Issues

We have shown that our algorithm avoids most of the visual artifacts as ghosting, aliasing and halving. However, our method may cause observable color shifts in the



Fig. 9. Comparison with average and blending pixel intensities. Averaging (left image) leads to ghosting, while blending intensities (right image) lead to visible transitions from day to night. Our method (Fig. 6) avoids both problems and maintains important gradients.

results. This phenomenon unfortunately has been a common problem of gradient-based approaches and can be observed in most previous works.<sup>5,6</sup> There are two major reasons that cause the color shifting. First of all, a valid vector field is not guaranteed to be maintained when modifying it with nonlinear operators. The gradient field of the result computed by our method is only an approximation of the desirable one. Secondly, in some cases, it is difficult to maintain the perception of high contrast in the result because the daytime and night-time snapshots are taken at significantly different exposure times.

Another issue is to capture the high-quality daytime background. Although we used medians of several images, in some cases some objects may remain in the frame for a long time. Good examples where this becomes an issue are the trucks parked on the ramp in Fig. 2 or the truck on the right lane in Fig. 8. Maintaining a running average or a database of daytime images may alleviate this problem.

A possible extension to our work will be to enforce the validity of the vector field when computing the gradients of the result. This requires using analytical operators to approximate our nonlinear mask and blending function. Separating intrinsic<sup>17</sup> and color images, then applying our algorithm on intrinsic images and fusing them back with the color images could be another possible solution.

## 6. Results

Our data for video enhancement is from the Washington State Department of Transportation website (used by permission). The data for enhancement using the basic algorithm was captured with a Canon PowerShot G3<sup>TM</sup> camera, placed on a fixed tripod. We show an example of a dynamic outdoor scene combined from a day and

a night snapshot (see Fig. 1). Notice the dark regions of the night-time snapshot are filled in by daytime snapshot pixels but with a smooth transition. We also show enhanced videos of traffic cameras (see Figs. 2 and 8). The camera resolution is  $320 \times 240$  pixels and it is very difficult to get an idea of the context, especially at night-time. In our experience, even on a well-organized website, where cameras are labeled and placed on a map, it is still hard to correctly evaluate the traffic situation because architectural features, which are essential for location recognition, cannot be readily discerned.

Processing was done offline as proof of concept and took approximately one second per frame after noise removal. We are working on a faster version of our method that can be applied to enhance traffic camera videos in real time.

### 6.1. User experience

One common complaint about techniques that create stylized outputs is the difficulty in judging their effectiveness and quantifying their usefulness. We performed an informal user study by asking seven users of various backgrounds to judge our results. Reactions to static enhanced night-time images were mixed. Some users at first were hesitant to believe the image is a (modified) photograph given its contradicting appearance: brightly lit buildings but night-time illumination of shops and streets. One user complained that the objects appeared deeper than they should be. Most users were, however, fascinated by the images. All users agreed that the enhanced images conveyed more information about the scene. Reactions to the enhanced videos were mostly positive when we asked which video they would like to see on the web for traffic cameras.

## 7. Conclusion

We have presented techniques to extract useful information from multiple snapshots taken using fixed cameras. By providing context to dark or low-quality snapshots or videos, we can create more useful images and easier to interpret surveillance videos. Our methods are suitable for processing low-contrast and noisy inputs while avoiding artifacts present in conventional combining methods such as aliasing, ghosting or halving.

## References

1. P. J. Burt and E. H. Adelson, A multiresolution spline with application to image mosaics, *ACM Trans. Graph.* **2**(4) (1983) 217–236.
2. Y. Chuang, B. Curless, D. Salesin and R. Szeliski, A Bayesian approach to digital matting, in *Proc. CVPR* **2** (2001) 264–271.
3. J. M. DiCarlo and B. A. Wandell, Rendering high dynamic range images, in *Proc. SPIE: Image Sensors* **3965** (2000), pp. 392–401.
4. F. Durand and J. Dorsey, Fast bilateral filtering for high-dynamic-range images, in *Proc. SIGGRAPH 2002*, ACM SIGGRAPH (2002), pp. 257–266.

5. R. Fattal, D. Lischinski and M. Werman, Gradient domain high dynamic range compression, in *Proc. SIGGRAPH 2002*, ACM SIGGRAPH (2002), pp. 249–256.
6. G. D. Finlayson, S. D. Hordley and M. S. Drew, Removing shadows from images, in *Proc. ECCV 4* (2002), pp. 823–836.
7. P. Haeberli, A multifocus method for controlling depth of field, Available at: <http://www.sgi.com/grafica/depth/index.html> (1994).
8. A. D. Ilie, R. Raskar and J. Yu, Gradient domain context enhancement for fixed cameras, in *Proc. ACCV 1* (2004), pp. 414–419.
9. S. B. Kang, M. Uyttendaele, S. Winder and R. Szelinski, High dynamic range video, in *Proc. SIGGRAPH 2003*, ACM SIGGRAPH (2003), pp. 319–325.
10. S. K. Nayar and S. G. Narasimhan, Vision in bad weather, in *Proc. ICCV* (1999), pp. 820–827.
11. P. Pèrez, M. Gagnat and A. Blake, Poisson image editing, in *Proc. SIGGRAPH 2003*, ACM SIGGRAPH (2003), pp. 313–318.
12. W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing* (Pearson Education, 1992).
13. E. Reinhard, M. Stark, P. Shirley and J. Ferwerda, Photographic tone reproduction for images, in *Proc. SIGGRAPH 2002*, ACM SIGGRAPH (2002), pp. 267–276.
14. S. M. Smith and J. M. Brady, SUSAN — a new approach to low level image processing, *Int. J. Comput. Vis.* **23**(1) (1997) 45–78.
15. D. Socolinsky and L. Wolff, A new visualization paradigm for multispectral imagery and data fusion, in *Proc. IEEE CVPR* (1999), pp. 319–324.
16. K. Toyama, J. Krumm, B. Brumitt and B. Meyers, Wallflower: principles and practice of background maintenance, in *Proc. ICCV* (1999), pp. 255–261.
17. Y. Weiss, Deriving intrinsic images from image sequences, in *Proc. ICCV 2* (2001), pp. 68–75.



**Adrian Ilie** is a Ph.D. candidate at the University of North Carolina at Chapel Hill. He received his M.S. degree in computer science from UNC and his B.S. in economic informatics from the Academy of Economic Studies in Bucharest, Romania.

His research interests span a range of topics in computer vision and computer graphics including camera network design for video surveillance, photometric calibration, and immersive display technologies.



**Ramesh Raskar** joined MERL as a Research Scientist in 2000 after his doctoral research at University of North Carolina at Chapel Hill, where he developed a framework for projector based displays. His work spans a range

of topics in computer vision and graphics including projective geometry, non-photorealistic rendering and intelligent user interfaces. He has developed algorithms for image projection on planar, non-planar and quadric curved surfaces that simplify constraints on conventional displays and has proposed Shader Lamps, a new approach for projector-based augmented reality.





**Jingyi Yu** is a Ph.D. candidate at Massachusetts Institute of Technology. He received his M.S. degree in electrical engineering and computer science from MIT and his B.S. in computer science and applied mathematics

from the California Institute of Technology.

His research interests span a range of topics in computer graphics and computer vision including image-based modeling and rendering, video surveillance, non-conventional optics and camera design, and graphics hardware.

Copyright of International Journal of Pattern Recognition & Artificial Intelligence is the property of World Scientific Publishing Company and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.