MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Image Fusion for Context Enhancement and Video Surrealism

R. Raskar, A. Ilie, J. Yu

TR2004-039 June 2004

Abstract

We present a class of image fusion techniques to automatically combine images of a scene captured under different illumination. Beyond providing digital tools for artists for creating surrealist images and videos, the methods can also be used for practical applications. For example, the non-realistic appearance can be used to enhance the context of nighttime traffic videos so that they are easier to understand. The context is automatically captured from a fixed camera and inserted from a day-time image of the same scene. Our approach is based on a gradient domain technique that preserves important local perceptual cues while avoiding traditional problems such as aliasing, ghosting and haloing. We presents several results in generating surrealistic videos and in increasing the information density of low quality nighttime videos.

International Symposium on Non-Photorealistic Animation and Rendering (NPAR)

Image Fusion for Context Enhancement and Video Surrealism

Ramesh Raskar*
Mitsubishi Electric Research Labs (MERL),†
Cambridge, USA

Adrian Ilie
UNC Chapel Hill, USA

Jingyi Yu
MIT, Cambridge, USA

Figure 1: Automatic context enhancement of a night time scene. The image is reconstructed from a gradient field. The gradient field is a linear blend of intensity gradients of the night time image and a corresponding day time image of the same scene.

Abstract

We present a class of image fusion techniques to automatically combine images of a scene captured under different illumination. Beyond providing digital tools for artists for creating surrealist images and videos, the methods can also be used for practical applications. For example, the non-realistic appearance can be used to enhance the context of nighttime traffic videos so that they are easier to understand. The context is automatically captured from a fixed camera and inserted from a day-time image (of the same scene). Our approach is based on a gradient domain technique that preserves important local perceptual cues while avoiding traditional problems such as aliasing, ghosting and haloing. We presents several results in generating surrealistic videos and in increasing the information density of low quality nighttime videos.

Keywords: image fusion, surrealism, gradient domain approach

1 Introduction

Nighttime images such as the one shown in Figure 1(a) are difficult to understand because they lack background context due to poor illumination. As a real life example, when you look at an image or video seen from a traffic camera posted on the web or shown on TV, it is very difficult to understand from which part of the town this image is taken, how many lanes the highway has or what buildings are nearby. All you see is pair of headlights moving on the screen (Fig 2(a)). How can we improve this image? Our solution is based on a very simple observation. We can exploit the fact that, the traffic camera can observe the scene all day long and create a high quality background. Then, we can simply enhance the context of the low quality image or video by fusing the appropriate pixels as shown in Figure 1(b) and 2(b) [Raskar et al. 2003b]. This idea appears to be very simple in retrospect. However, despite our search efforts, this concept of image fusion appears to have been unexplored. The closest work we found of combining daytime and nighttime images of the same scene came from a rather unexpected source: a Surrealist painting by René Magritte.

Surrealism is the practice of producing incongruous imagery by means of unnatural juxtapositions and combinations [Merriam-Webster 2001]. In the well-known surrealist painting by Rene Magritte's , 'The Empire of Lights', a dark, nocturnal street scene is set against a pastel-blue, light-drenched sky spotted with fluffy cumulus clouds, with no fantastic element other than the single paradoxical combination of day and night. Each part on its own looks real, but it is the fusion of parts that gives a strange non-realistic appearance in the overall context (Figure 3). Inspired by this notion of adding unusual context, in this paper we present a class of image fusion techniques to automatically blend different images of the same scene into a seamless rendering. We are not artists and computer generated fusion is unlikely to evoke similar emotions. But we hope to provide digital tools for artists to create new types of surrealist images as well as videos.

1.1 Overview

Our image fusion approach is based on a gradient domain technique that preserves important local perceptual cues while avoiding tradi-

*raskar@merl.com, adyilie@cs.unc.edu, jingyi@graphics.csail.mit.edu
†http://www.merl.com/projects/NPRfusion/

Figure 2: Enhanced traffic video. A low quality nighttime image, and a frame from the final output of our algorithm.

tional problems such as aliasing, ghosting and haloing. We first encode the pixel importance based on local variance in input images or videos. Then, instead of a convex combination of pixel intensities, we use linear combination of the intensity gradients where the weights are scaled by the pixel importance. The image reconstructed from integration of the gradients achieves a smooth blend of the input images, and at the same time preserves their important features.

1.2 Related Work

Fusion We can classify the methods to combine information from multiple images into one by noting which parameter of the scene or the camera is changing between successive images. The main idea can be traced to the 19th century by Marey and Murbridge [Muybridge 1985; Braun 1992], who created beautiful depictions of objects over time. Fusing images varying in depth include the classic depiction of motion and shape in Duchamps Nude Descending a Staircase. This has been extended by Freeman and Zhang [2003] via a stereo camera. The automatic method is called shapetime photography and also explored by [Essa 2002]. Images captured by varying camera exposure parameters are used to generate high-dynamic range (HDR) images. Tone mapping for compression of such images includes gradient space [Fattal et al. 2002] and image space [Durand and Dorsey 2002; Reinhard et al. 2002] methods. Images captured with varying focus can also be combined to create all-in-focus imagery [Haeberli 1994]). Images with varying camera viewpoint are combined in cubism, using multiple-center-of-projection images [Rademacher and Bishop 1998] and by reflectance map of laser scanners. In this sense, our method can be considered as fusion of images varying in natural illumination.

Novel images have been created via video cubes, i.e. by slicing 3D volume of pixels, by [Fels and Mase 1999; Klein et al. 2002b; Cohen 2003]. We explore non-planar cuts in the cubes of spatio-temporal gradient fields of videos.

Videos When stylization and non-photorealistic rendering (NPR) methods designed for static images are applied to video sequences on a frame-by-frame basis, the results generally contain undesirable temporal artifacts. To overcome these artifacts, [Meier 1996; Litwinowicz 1997; Hertzmann and Perlin 2000] used optical flow information to maintain temporal coherency. We use a similar approach in the gradient domain.

Interesting spatial composition of multiple videos is created for art via video mosaics [Klein et al. 2002a] or for shot composition and overlay rules, based on 'region objects' [Gleicher et al. 2002] in the context of virtual videography of classroom teaching. Our method of achieving temporal coherency is related to the region objects which are defined by pixel neighborhoods in space and time. Sophisticated image and video matting methods [Chuang et al. 2001] are also useful for foreground segmentation.

Gradient domain methods Our approach is inspired by some recent methods that work in the gradient space rather than intensity space.

Image conflation and fusion of multi-spectral imagery to merge satellite imagery captured at different wavelengths is a common application [Socolinsky and Wolff 1999]. The images are relatively similar and the output is not always rendered in a pseudo-photorealistic manner.

For compressing HDR images [Fattal et al. 2002] attenuate large image gradients before image reconstruction. However, our problem is different from combining high dynamic range images. In HDR images, the pixel intensities in successive images increase monotonically allowing one to build a single floating point format image. This is not possible in our case. In day-night images we see intensity gradient reversals (such as objects that are darker than their surroundings during the day, but brighter than their surroundings during the night).

Pèrez et al. [Pèrez et al. 2003] present a useful interactive image editing technique that uses integration of modified gradients to support seamless pasting and cloning. However, since their goal is to provide a framework for image editing, they rely on user input to manually assign the region from which the gradients are taken.

Authors of [Finlayson et al. 2002] remove 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. [Weiss 2001] uses a similar method for creating intrinsic images to reduce shadows.

Figure 3: The Empire of Light, by René Magritte (Used by permission, Guggenheim Museum, New York).

1.3 Contributions

Our main contribution is the idea of exploiting information available from fixed cameras to create context-rich images. Our technical contributions include the following.

- A scheme for asymmetrically fusing multiple images preserving useful features to improve the information density in a picture;
- A method for temporally-coherent context enhancement of videos in presence of unreliable frame differencing.

In addition, we modify the method of image reconstruction from gradients fields to handle the boundary conditions to overcome integration artifacts. We employ a color assignment strategy to re-

duce the commonly known artifact of the gradient-based method –observable color shifting.

A fused image should be "visually pleasing", i.e., it should have very few aliasing, ghosting or haloing artifacts and it should maintain smooth transition from background to foreground. Our method achieves 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.

Our proposed algorithm consists of two major steps similar to video matting: foreground extraction and background fusion. Robust foreground extraction in image space is difficult to achieve in practice, especially when dealing with low contrast and noisy images and videos. Therefore we propose a gradient space algorithm.

A gradient space method also allows us to simply state the constraints on the resultant image i.e. which parts of constituent images should be preserved. Then we search for an optimal image that satisfies gradient image in the least-square error sense. Compared to gradient domain approaches described above, our approach attempts to fuse images that are sufficiently different. We are inspired by many of the techniques mentioned above and aim to address some of their limitations. We support automatic region selection, allow linear blend of gradient fields and extend the technique to support video synthesis.

2 Image Fusion

We present the basic algorithm for image fusion followed by our approach to ensure better image reconstruction and color assignment.

2.1 Basic Algorithm

How would one combine information from two (or more) images in a meaningful way? How would one pick high-quality background parts from a daytime image while keeping all the low-quality important information from a nighttime image? The traditional approach is to use a linear combination of the input images. We instead specify the desired local attributes of the final image and solve the inverse problem of obtaining a global solution that satisfies the local attributes. This leads to a non-linear combination, which means pixels with the same intensities map to different intensities in the final image. Our basic idea for determining the important areas of each image relies on the widely accepted assumptions [DiCarlo and Wandell 2000] that the human visual system is not very sensitive to absolute luminance reaching the retina, but rather responds to local intensity ratio changes. Hence, the local attribute is the local variance and we define an importance function for each input image based on the spatial and temporal intensity gradients, which are a measure of the local spatial and temporal variance.

Our approach is based on two heuristics. (a) We carry into the desired image the gradients from each input image that appear to be locally important and (b) we provide context to locally-important areas while maintaining intra-image coherence. Note that we do not improve the quality of the pixels themselves, but simply give sufficient context to improve human interpretation. Hence any operations such as contrast enhancement, histogram equalization, mixed Gaussian models for background estimation [Toyama et al. 1999] are orthogonal to our approach and can be easily used alongside to improve the final result.

The regions of high spatial variance across one image are computed by thresholding the intensity gradients, $G = (G^X, G^Y)$, for the horizontal and vertical directions using a simple forward difference. The regions of high temporal variance between two images are computed by comparing the intensity gradients of corresponding pixels from the two images. We then compute an *importance image* (a weighting function) W, by processing the gradient magnitude $|G|$. The weighted combination of input gradients gives us the gradient of the desired output (Figure 4). The basic steps are described in Algorithm 1.

Algorithm 1 Basic algorithm

for each input image I_i **do**
  Find gradient field $G_i = \nabla I_i$
  Compute importance image W_i from $|G_i|$
end for
for each pixel (x,y) **do**
  Compute mixed gradient field $G(x,y) = \sum_i W_i(x,y)G_i(x,y) / \sum_i W_i(x,y)$
end for
Reconstruct image I' from gradient field G
Normalize pixel intensities in I' to closely match $\sum_i W_i I_i$

As described in the following sections, the process of determining importance weights W_i, depends on the specific application.

2.2 Image Reconstruction

Image reconstruction from gradients fields, an approximate invertibility problem, is 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 [Fattal et al. 2002] to minimize $|\nabla I' - G|$. The estimate of the desired intensity function I', so that $G = \nabla I'$, can be obtained by solving the Poisson differential equation $\nabla^2 I' = divG$, involving a Laplace and a divergence operator. We use the full multigrid method [Press et al. 1992] to solve the Laplace equation. We pad the 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.

2.3 Boundary Conditions

One needs to specify boundary conditions to solve the Laplace equation (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. To reduce image artifacts near the boundary, we modify the source image, I, by padding it with colors obtained by Gaussian smoothing boundary pixels. The reconstructed image is later cropped to the original size. Padding by 5 pixels was found sufficient. Figure 5 shows a comparison of integrating the gradient field of an image with and without padding.

Figure 5: Overcoming integration artifacts by padding. Is it possible to recover an image from its gradient ? (Left to right) The original image, the integration of gradient of original image with padding, and without padding.

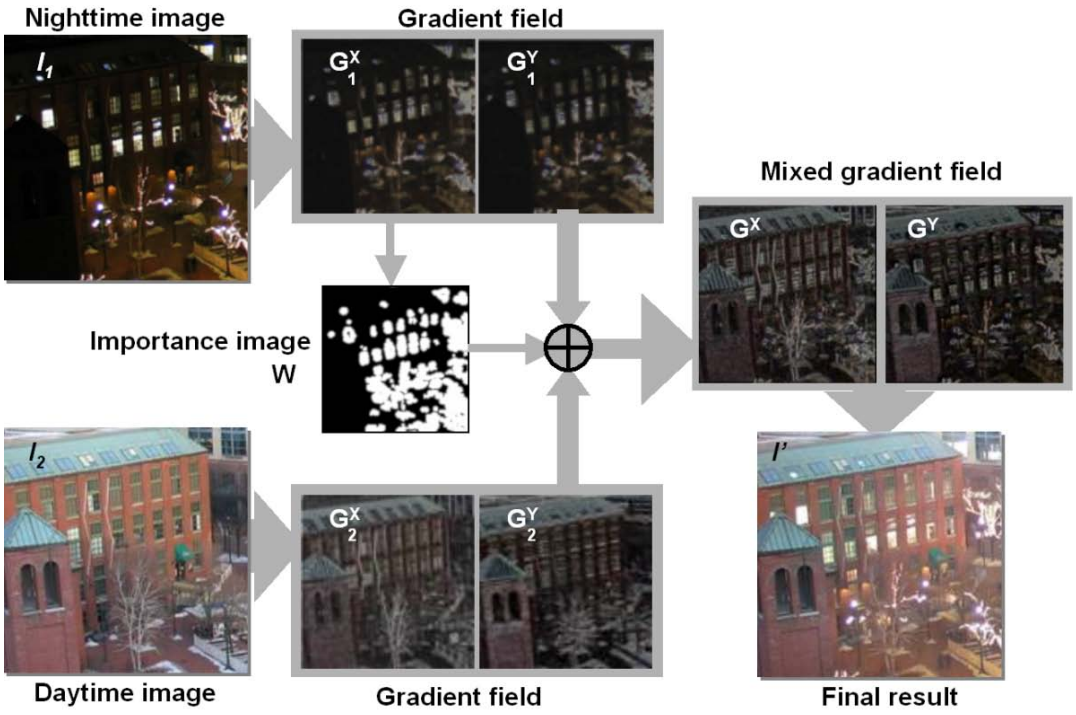


Figure 4: Flowchart for asymmetric fusion. Importance image is derived from only the night time image. Mixed gradient field is created by linearly blending intensity gradients.

2.4 Color Assignment

Before we obtain the final image, I'', it is important to note that the pseudo-integration of the gradient field involves a scale and shift ambiguity, $I''(x,y) = c_1 I'(x,y) + c_2$. We compute the unknowns, c_1 and c_2, (in the least square sense) using a simple heuristics that the overall appearance of each part of the reconstructed image should be close to the corresponding part of the foreground and background images. Each pixel leads to a linear equation, $\sum W_i I_i(x,y) = c_1 I'(x,y) + c_2$. We do image reconstruction in all three color channels separately and compute the unknowns per channel. Thanks to the boundary condition $(\nabla I' \cdot n = 0)$, the scale and shift in the three channels do not introduce noticeable artifacts.

3 Context Enhancement of Images

We build our results on the basic observation that if the camera and viewed geometry remain static, only illumination and minor parts of the scene change (e.g., moving objects like people, devices, vehicles). Thus, the intensity gradients corresponding to the stationary parts in the poor-context night image can be replaced with better quality image gradients from a high-contrast day image.

Static scenes: When the scene geometry is the same and only the illumination changes, the context can clarify the scene and help identify areas of interest. Imagine trying to capture the view from a distance of Times Square in New York at daytime and nighttime within a single picture. This may be used in tourism brochures, for advertising, art or for simple visualization.

Dynamic scenes: More interesting applications are when there is a change in scene geometry. Using the notions of a static background and a dynamic foreground, we can provide context for an action or event. The dynamic component can be captured in multiple snapshots or in a video. One example are surveillance videos, where context can help answering questions such as: why is a person standing near a part of a building (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). Another example 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 an image captured using brighter illumination, when there are no visitors in the scene. Finally, using a higher resolution background image can increase the perceived resolution of the dynamic foreground.

3.1 Enhancement of Static Scenes

A simple choice, used by the authors of [Pèrez et al. 2003], is to use desired gradient field as the local maximum of all input gradients, $G(x,y) = max_i(G_i(x,y))$. In this case importance weights are either 0 or 1. A better choice, in our case, is to give more importance to nighttime gradients in region of the nighttime image where gradients or intensities are above a fixed threshold. This is to make sure that no information in the nighttime image is lost in the final image. Additionally, user input can help guide the algorithm by manually modifying the importance image.

3.2 Enhancement of Dynamic Scenes

When dealing with dynamic scenes, the goal is to provide context to the foreground changes in the night image by replacing low-detail background areas. This is where many of the traditional method using linear combination will fail to create seamless images. Let us consider the case where we want to provide context to nighttime image N using information from another nighttime reference image R and a daytime image D. We create a new mask image M,

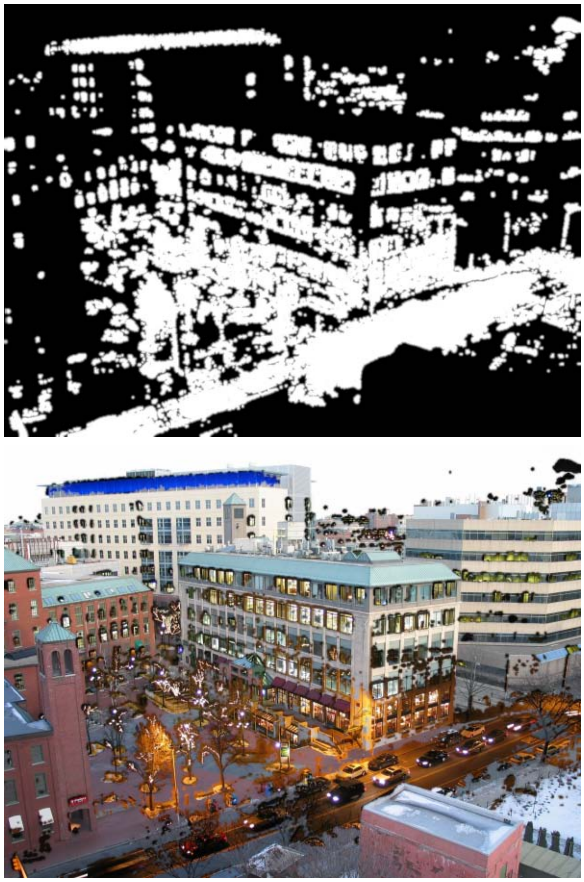


Figure 6: Importance mask used for blending gradients and Result of simple linear blend of pixel intensities showing artifacts.

and set $M(x, y) = |N(x, y) - R(x, y)|$ so that the importance is scaled by the difference between the two nighttime images. Mask M is thresholded and normalized, then multiplied by the weights for image N. (See Figure 6 top and 8)

Although we use a very simple segmentation technique (pixel-wise difference in color space between images 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. 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 images, 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 image segmentation techniques would bring marginal improvements to our results.

4 Context Enhancement of Video

Providing context to captured events and actions can also enhance low quality videos, such as the ones obtained from security and traffic surveillance cameras. The context, as in the previous subsection, comes from a single higher-quality day image. Videos, however, present several additional challenges: (a) inter-frame coherence must also be maintained i.e. the weights in successive images should change smoothly and (b) a pixel from a low quality image may be important even if the local variance is small (e.g., the area between the headlights and the taillights of a moving car).

Figure 7: Day Image.

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 weight high in the interval between the head and the tail. The steps are described in Algorithm 2.

Algorithm 2 Context enhancement of video

```
Compute spatial gradients of daytime image D = ∇I
Smooth video
for each video frame F_j do
  Compute spatial gradients N_j = ∇F_j
  Find binary masks M_j by thresholding temporal differences
  Create weights for temporal coherence W_j using M_j
  for each pixel (x,y) do
    if W_j(x,y) > 0 then
      Compute mixed gradient field as G(x,y) = N_j(x,y) *
      W_j(x,y) + D * (1 − W_j(x,y))
    else
      Compute mixed gradient field as G(x,y) =
      max(|D(x,y)|, |N_j(x,y)|)
    end if
  end for
  Reconstruct frame F'_j from gradient field G
  Normalize pixel intensities in F'_j to closely match F_j(x,y) *
  W_j(x,y) + I * (1 − W_j(x,y))
end for
```

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_j is calculated by thresholding the difference with the previous frame, $|F_j - F_{j-1}|$. 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=j-c..j+c. We chose the extent of influence c, to be 5 frames in each direction. Thus, the weight due to temporal variation W_j is a mask with values in [0,1] that vary smoothly in space and time. Then for each pixel of each frame, if $W_j(x, y)$ is non-zero, we use the method of context enhancement of dynamic scene i.e. blend the gradients of the night frame and day frame scaled by W_j and $(1 - W_j)$. If $W_j(x, y)$ is zero, we revert to a special case of the method of enhancement for static scenes. Finally, each frame is individually reconstructed from the mixed gradient field for that frame. (See Figure 9).

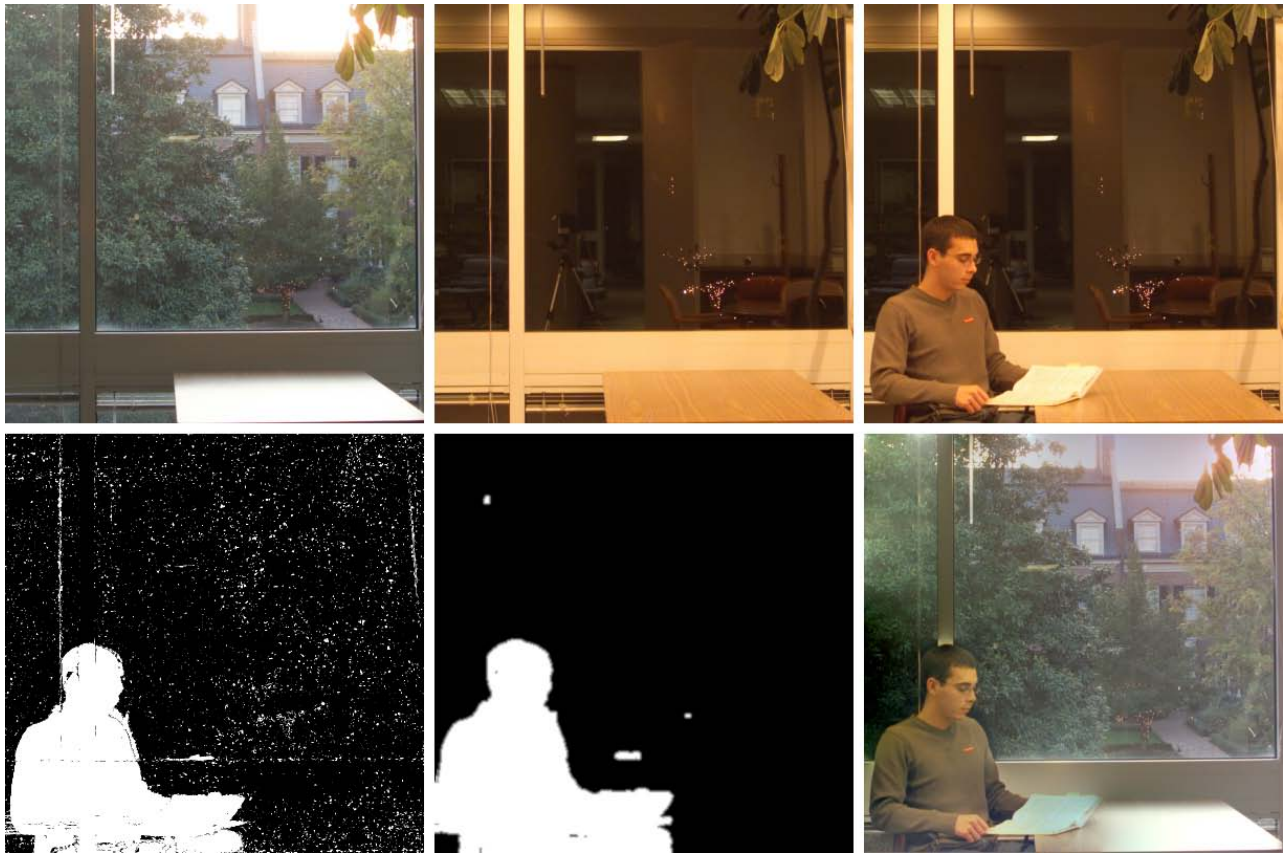


Figure 8: Enhancing a dynamic scene. (Top row) A high quality daytime image, a nighttime reference, and with a foreground person, (Bottom row) A simple binary mask obtained by subtracting reference from foreground, the importance image obtained after processing the binary mask, and the final output of our algorithm.

The input video is noise reduced by using 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 [Smith and Brady 1997] (3x3x5 neighborhood, sigma=15 and t=20). The high-quality daytime image used for filling in the context is obtained by median filtering a daytime video clip (about 15 seconds).

Just as in the case of images, a good quality video segmentation or optical flow technique will 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 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 Stylization

We show how the gradient space method can be used as a tool to create artistic, surrealist effects. We demonstrate procedures to transform time-lapse images taken over a whole day into a single image or a video [Raskar et al. 2003a]. We are not artists, so these are our feeble attempts at exploiting the mechanism to give the reader an idea of the kind of effects possible. It should be understood that a skilled artist can harness the power of our tools to create compelling visual effects.

5.1 Mosaics

Given an image sequence of views of the same scene over time, we are interested in cuts of the video cube that represent the overall scene. Here we present some potential opportunities. Arbitrary planar slices of video cubes are also interesting ([Fels and Mase 1999; Klein et al. 2002b; Cohen 2003]). However, Our aim to preserve the overall appearance of the scene.

Consider a specific case of creating a mosaic of vertical strips. A simple method would involve a diagonal cut in the video cube, i.e. each column of the final image coming from the corresponding column in successive images (Figure 10.left). However, we need to address several issues. First, since the time-lapse images are taken every few minutes, the temporal sampling is usually not very dense. Hence, the fused strips look discontinuous. Second, most of the dramatic changes happen in a very short time e.g. around sunrise or sunset. So the sampling is required to be non-linear with denser samples at sunrise and sunset. Third, we would like to maintain visual seamlessness across strips. An obvious choice is partial overlap and blending of strips.

We instead blend the intensity gradients of each vertical strip. By integrating the resultant gradient field, we ensure a smooth synthesis that preserves the overall appearance but allows smooth variation in illumination (Figure 10.right). It would be interesting to add more abstraction as a postprocess on fused images [DeCarlo and Santella 2002].

Figure 10: Stylization by mosaicing vertical strips of a day to night sequence (Left) Naive algorithm (Right)The output of our algorithm.

Figure 9: Enhancing traffic video. (Top row) A high quality daytime and a low quality nighttime image, (Bottom row) The importance image obtained after processing, and the final output of our algorithm. Notice the road features and background buildings (Video available at project website).

5.2 Videos

In the spirit of the surrealist painting by Magritte, we aim to create a surrealistic video where a night and a day event is visible at the same time. First, in a straightforward reversal of our night-video enhancement method, we fuse a each frame of a day time sequence with a fixed frame of a nighttime image. In the accompanying video we show how daytime shadows appear to move in a night time scene. One other possibility is the fusion of two images taken at different times during the day. This can be extended to fusing successive frames of two input videos into a non-realistic output video. In the accompanying video we show how shadows and street lights mix to create an unusual fusion of natural and artificial lighting.

Finally, consider a video fusion where, the pair being fused have a different temporal rate of sampling. In the accompanying video, a sunrise sequence is fused with a sunset sequence creating an illusion of a bi-solar illumination. However, the shadows at sunset move much quicker than the rate at which the sun appears to be rising.

6 Discussion

6.1 Comparison

A naïve approach to automatically combining a daytime and nighttime picture would be to use a pure pixel substitution method based on some importance measure. This works well only when the source images are almost identical (e.g. two images of the same scene with different focus [Haeberli 1994]). Similarly, blending strategies such as $max_i(I_i(x,y))$ or $average_i(I_i(x,y))$ also create problems. For example, when combining day-night images, one needs to deal with high variance in daytime images and with mostly low contrast and patches of high contrast in night images. Taking the average simply overwhelms the subtle details in the nighttime image, and presents 'ghosting' artifacts around areas that are bright at nighttime. 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 images. Figure 11 shows a comparison between averaging pixel values, blending pixel values using an importance function, and our method.

6.2 Issues

We have shown that our algorithm avoids most of the visual artifacts as ghosting, aliasing and haloing. However our method may cause observable color shifts in the resulting images, especially when the segmented foreground occupies a substantially large portion in the result. This phenomenon unfortunately has been a common problem of gradient-based approaches and can be observed in most previous works [Finlayson et al. 2002], [Fattal et al. 2002]. 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 non-linear operators. The gradient field of the resulting image 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 a single image because the day and night time images are captured at significantly different exposure times.

A minor but important issue is capturing of the high-quality background. Although we used medians of several images, in some cases some object may remain in the frame for a long time. A good

Figure 11: Comparison with average and blending pixel intensities. using weights. Averaging (left image) leads to ghosting, while blending intensities (right image) leads to visible transitions from day to night. Our method (Figure 8) avoids both problems and maintains important gradients. But in order to maintain seamlessness in images with different white balance, it introduces color shifts and bleeding.

example where this becomes an issue is the trucks parked on the ramp in Figure 2.

A possible extension to our work will be to maintain a valid vector field when modifying the gradient image. This requires using analytical operators to approximate our non-linear mask and blending function. This remains an active area of research and we hope to use better reconstruction algorithms in the future. Separating intrinsic [Weiss 2001] and color images, then applying our algorithm on intrinsic images and fusing them back with the color images could be another possible solution.

7 Results

Our data for video enhancement is from the Washington State Dept. of Transportation website (used by permission). The data for image enhancement was captured with a Canon PowerShot G3 camera, placed on a fixed tripod.

We show an example of outdoor scene combined from a day and a night picture (see Figure 1). Notice the dark regions of the night image are filled in by day image pixels but with a smooth transition. We also show enhanced videos of traffic cameras (see Figure 2). The camera resolution is 320x240 pixels and it is very difficult to get an idea of the context, especially at nighttime. While this type of images is usually enough for a trained traffic controller, if one is not familiar with location, showing a nighttime traffic image makes it very difficult to understand where the lanes and exits on the highway are. In our experience, even on a well-organized website, where cameras are labeled and placed on a map, it is still difficult to correctly evaluate the traffic situation because it is difficult to discern architectural features in the image, which are essential for location recognition.

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 images in real time. One advantage of our method is that there are very few parameters to be tuned, allowing us to use a simple interface to enhance different videos.

7.1 User Experience

One common complaint about techniques that create stylized outputs is the difficulty in judging their effectiveness. We do not dare to compare of our computer generate art to paintings by artists, but we performed an informal user study by asking 7 users of various backgrounds to judge our results in terms of usefulness. Reactions to static enhanced nighttime images were mixed. Some users at first were hesitant to believe the images are (modified) photographs given their contradicting appearances: brightly lit buildings but nighttime 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. One called it 'dreamy', an interesting comment recalling the surrealist Magritte painting. 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. It was considered better than the current method of traffic camera websites which show live shot alongside a sample day-time shot, side-by-side [BostonWebcams 2003]. Of course this option is not available when the video is being show on TV's at home.

Some complaints included the perception of fog or rain. This is probably due to the image of glaring car headlights in a day-like setting. However, the images do give a wrong impression about the scene or the weather if users are not informed.

Based on feedback, video security experts do not want enhanced videos when they are actively observing the data. They may switch back and forth between original and enhanced view to get the context. However, apparently, when a agent is looking at rows of multiple TVs, it is difficult for even a well-trained agent to remember which view corresponds to which camera. We were told that, in a passive mode of observation, it may be ok to leave the enhanced views on so that the agent can orient him/herself with a quick glance.

8 Conclusion

We have presented gradient domain techniques to extract useful information from multiple images of scenes by exploiting their illumination-dependent properties. Our methods are useful as digital tools for artists to create surrealist images and videos. By providing context to dark or low quality images or videos, we can create more comprehensible images and generate information rich video streams.

Acknowledgements We would like to thank Hongcheng Wang for useful discussions and Fredo Durand for suggesting the comparison with Magritte painting.

Project website for videos http://www.merl.com/projects/NPRfusion/.

References

BOSTONWEBCAMS, 2003. (Observe the cameras at night). http://www.boston.com/traffic/cameras/artery.htm.

BRAUN, M. 1992. *Picturing Time*. University of Chicago.

CHUANG, Y., CURLESS, B., SALESIN, D., AND R., S. 2001. A Bayesian Approach to Digital Matting. In *Proceedings of CVPR*, vol. 2, 264–271.

COHEN, M., 2003. Image Stacks. Presentation at MIT, Cambridge, USA, October 2003.

DECARLO, D., AND SANTELLA, A. 2002. Stylization and Abstraction of Photographs. In *Proc. Siggraph 02, ACM Press*.

DICARLO, J., AND WANDELL, B. 2000. Rendering High Dynamic Range Images. In *Proceedings of SPIE: Image Sensors*, vol. 3965, 392–401.

DURAND, F., AND DORSEY, J. 2002. Fast Bilateral Filtering for High-Dynamic-Range Images. In *Proceedings of SIGGRAPH 2002*, ACM SIGGRAPH, 257–266.

ESSA, I., 2002. Graphical Display of Motion-Capture System using a Shape-time Style Rendering.

FATTAL, R., LISCHINSKI, D., AND WERMAN, M. 2002. Gradient Domain High Dynamic Range Compression. In *Proceedings of SIGGRAPH 2002*, ACM SIGGRAPH, 249–256.

FELS, S., AND MASE, K. 1999. Interactive video cubism. In *Workshop on New Paradigms in Information Visualization and Manipulation*, ACM Press, 78–82.

FINLAYSON, G., HORDLEY, S., AND DREW, M. 2002. Removing Shadows from Images. In *Proceedings of ECCV*, vol. 4, 823–836.

FREEMAN, B., AND ZHANG, H. 2003. Shapetime photography. In *Proceedings of CVPR*, vol. 2, 264–271.

GLEICHER, M., HECK, R., AND WALLICK, M. 2002. A framework for virtual videography. In *Smart Graphics*.

HAEBERLI, P., 1994. A Multifocus Method for Controlling Depth of Field. Availavle at: http://www.sgi.com/grafica/depth/index.html.

HERTZMANN, A., AND PERLIN, K. 2000. Painterly Rendering for Video and Interaction. In *Proceedings of NPAR 2000, Symposium on Non-Photorealistic Animation and Rendering (Annecy, France, June 2000)*, ACM, 7–12.

KLEIN, A. W., GRANT, T., FINKELSTEIN, A., AND COHEN, M. F. 2002. Video Mosaics. In *Proceedings of NPAR 2002, International Symposium on Non Photorealistic Animation and Rendering (Annecy, France, June2002)*.

KLEIN, A. W., SLOAN, P.-P. J., FINKELSTEIN, A., AND COHEN, M. F. 2002. Stylized Video Cubes. In *ACM SIGGRAPH Symposium on Computer Animation*, 15–22.

LITWINOWICZ, P. 1997. Processing Images and Video for an Impressionist Effect. In *Proceedings of SIGGRAPH'97 (Los Angeles, Aug. 1997)*, T. Whitted, Ed., Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH, 407–414.

MEIER, B. J. 1996. Painterly Rendering for Animation. In *Proceedings of SIGGRAPH'96 (New Orleans, Aug. 1996)*, H. Rushmeier, Ed., Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH, 477–484.

MERRIAM-WEBSTER. 2001. *Collegiate Dictionary*.

MUYBRIDGE, E. 1985. *Horses and other animals in motion*. Dover.

PÈREZ, P., GANGNET, M., AND BLAKE, A. 2003. Poisson image editing. In *Proceedings of SIGGRAPH 2003*, 313–318.

PRESS, W. H., TEUKOLSKY, S., VETTERLING, W. T., AND FLANNERY, B. P. 1992. *Numerical Recipes in C: The Art of Scientific Computing* . Pearson Education.

RADEMACHER, P., AND BISHOP, G. 1998. Multiple-Center-of-Projection Images. In *Proceedings of SIGGRAPH 98 (Orlando, July 1998)*, M. Cohen, Ed., Annual Conference Series, ACM SIGGRAPH, 199–206.

RASKAR, R., YU, J., AND ILIE, A., 2003. Stylized Images using Variable Illumination. Unpublished, January 2003.

RASKAR, R., YU, J., AND ILIE, A., 2003. Video Surveillance with NPR Image Fusion. http://www.merl.com/projects/NPRfusion/, March 2003.

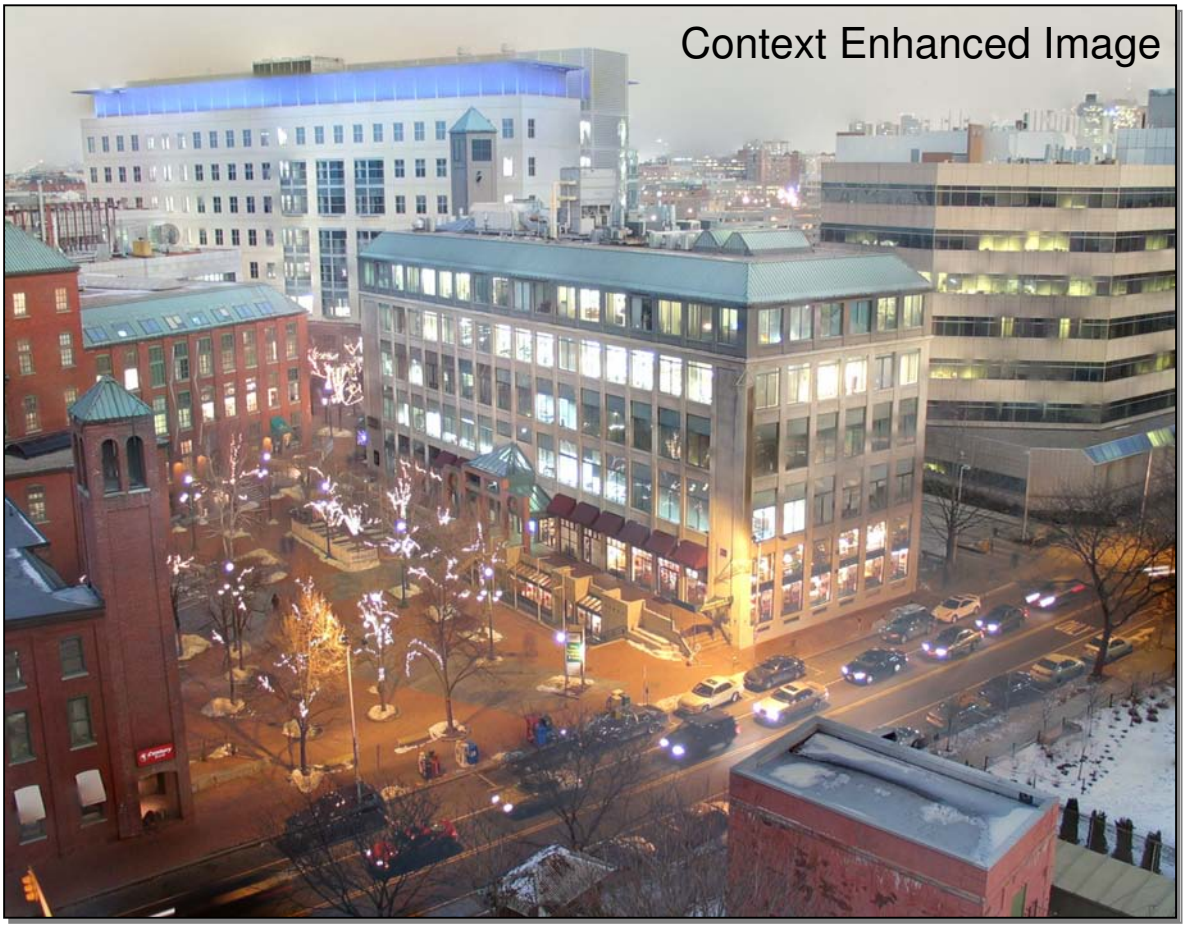
REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J. 2002. Photographic Tone Reproduction for Images. In *Proceedings of SIGGRAPH 2002*, ACM SIGGRAPH, 267–276.

SMITH, S., AND BRADY, J. 1997. SUSAN - a new approach to low level image processing. *Int. Journal of Computer Vision 23*, 1, 45–78.

SOCOLINSKY, D., AND WOLFF, L. 1999. A New Visualization Paradigm for Multispectral Imagery and Data Fusion. In *Proceedings of IEEE CVPR*, 319–324.

TOYAMA, K., KRUMM, J., BRUMITT, B., AND MEYERS, B. 1999. Wallflower: Principles and Practice of Background Maintenance. In *ICCV*, 255–261.

WEISS, Y. 2001. Deriving Intrinsic Images From Image Sequences. In *Proceedings of ICCV*, vol. 2, 68–75.


Context Enhanced Image


Night-time Image


Importance Image


Day-time Image


Pixel-blending Results