

Seamless Image Stitching in the Gradient Domain^{*}

Anat Levin, Assaf Zomet^{**}, Shmuel Peleg, and Yair Weiss

School of Computer Science and Engineering

The Hebrew University of Jerusalem

91904, Jerusalem, Israel

{alevin,peleg,yweiss}@cs.huji.ac.il, zomet@cs.columbia.edu

Abstract. Image stitching is used to combine several individual images having some overlap into a composite image. The quality of image stitching is measured by the similarity of the stitched image to each of the input images, and by the visibility of the seam between the stitched images.

In order to define and get the best possible stitching, we introduce several formal cost functions for the evaluation of the quality of stitching. In these cost functions, the similarity to the input images and the visibility of the seam are defined in the gradient domain, minimizing the disturbing edges along the seam. A good image stitching will optimize these cost functions, overcoming both photometric inconsistencies and geometric misalignments between the stitched images.

This approach is demonstrated in the generation of panoramic images and in object blending. Comparisons with existing methods show the benefits of optimizing the measures in the gradient domain.

1 Introduction

Image stitching is a common practice in the generation of panoramic images and applications such as object insertion, super resolution [1] and texture synthesis [2]. An example of image stitching is shown in Figure 1. Two images I_1, I_2 capture different portions of the same scene, with an overlap region viewed in both images. The images should be stitched to generate a mosaic image I . A simple pasting of a left region from I_1 and a right region from I_2 produces visible artificial edges in the seam between the images, due to differences in camera gain, scene illumination or geometrical misalignments.

The aim of a stitching algorithm is to produce a visually plausible mosaic with two desirable properties: First, the mosaic should be as *similar* as possible to the input images, both geometrically and photometrically. Second, the seam between the stitched images should be *invisible*. While these requirements are

^{*} This research was supported (in part) by the EU under the Presence Initiative through contract IST-2001-39184 BENOGO.

^{**} Current Address: Computer Science Department, Columbia University, 500 West 120th Street, New York, NY 10027

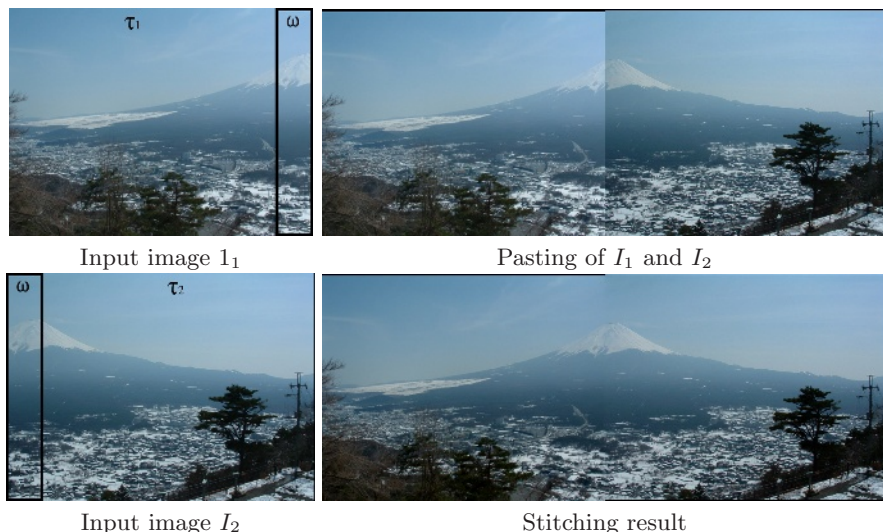


Fig. 1. Image stitching. On the left are the input images. ω is the overlap region. On top right is a simple pasting of the input images. On the bottom right is the result of the GIST1 algorithm.

widely acceptable for visual examination of a stitching result, their definition as quality criteria was either limited or implicit in previous approaches.

In this work we present several cost functions for these requirements, and define the mosaic image as their optimum. The stitching quality in the seam region is measured in the gradient domain. The mosaic image should contain a minimal amount of seam artifacts, i.e. a seam should not introduce a new edge that does not appear in either I_1 or I_2 . As image dissimilarity, the gradients of the mosaic image I are compared with the gradients of I_1, I_2 . This reduces the effects caused by global inconsistencies between the stitched images. We call our framework GIST: Gradient-domain Image STitching.

We demonstrate this approach in panoramic mosaicing and object blending. Analytical and experimental comparisons of our approach to existing methods show the benefits in working in the gradient domain, and in directly minimizing gradient artifacts.

1.1 Related Work

There are two main approaches to image stitching in the literature, assuming that the images have already been aligned. Optimal seam algorithms[3,2,4] search for a curve in the overlap region on which the differences between I_1, I_2 are minimal. Then each image is copied to the corresponding side of the seam. In case the difference between I_1, I_2 on the curve is zero, no seam gradients are produced in the mosaic image I . However, the seam is visible when there is no such curve,

for example when there is a global intensity difference between the images. This is illustrated on the first row of Figure 2. In addition, optimal seam methods are less appropriate when thin strips are taken from the input images, as in the case of manifold mosaicing [5].

The second approach minimizes seam artifacts by smoothing the transition between the images. In Feathering [6] or alpha blending, the mosaic image I is a weighted combination of the input images I_1, I_2 . The weighting coefficients (alpha mask) vary as a function of the distance from the seam. In pyramid blending[7], different frequency bands are combined with different alpha masks. Lower frequencies are mixed over a wide region, and fine details are mixed in a narrow region. This produces gradual transition in lower frequencies, while reducing edge duplications in textured regions. A related approach was suggested in [8], where a smooth function was added to the input images to force a consistency between the images in the seam curve. In case there are misalignments between the images[6], these methods leave artifacts in the mosaic such as double edges, as shown in Figure 2.

In our approach we compute the mosaic image I by an optimization process that uses image gradients. Computation in the gradient domain was recently used in compression of dynamic range[9], image editing [10], image inpainting [11] and separation of images to layers [12,13,14,15]. The closest work to ours was done by Perez et. al. [10], who suggest to edit images by manipulating their gradients. One application is object insertion, where an object is cut from an image, and inserted to a new background image. The insertion is done by optimizing over the derivatives of the inserted object, with the boundary determined by the background image. In sections 4, 5 we compare our approach to [10].

2 GIST: Image Stitching in the Gradient Domain

We describe two approaches to image stitching in the gradient domain. Section 2.1 describes GIST1, where the mosaic image is inferred directly from the derivatives of the input images. Section 2.2 describes GIST2, a two-steps approach to image stitching. Section 2.3 compares the two approaches to each other, and with other methods.

2.1 GIST1: Optimizing a Cost Function over Image Derivatives

The first approach, GIST1, computes the stitched image by minimizing a cost function E_p . E_p is a dissimilarity measure between the derivatives of the stitched image and the derivatives of the input images.

Specifically, let I_1, I_2 be two aligned input images. Let τ_1 (τ_2 resp.) be the region viewed exclusively in image I_1 (I_2 resp.), and let ω be the overlap region, as shown in Figure 1, with $\tau_1 \cap \tau_2 = \tau_1 \cap \omega = \tau_2 \cap \omega = \emptyset$. Let W be a weighting mask image.

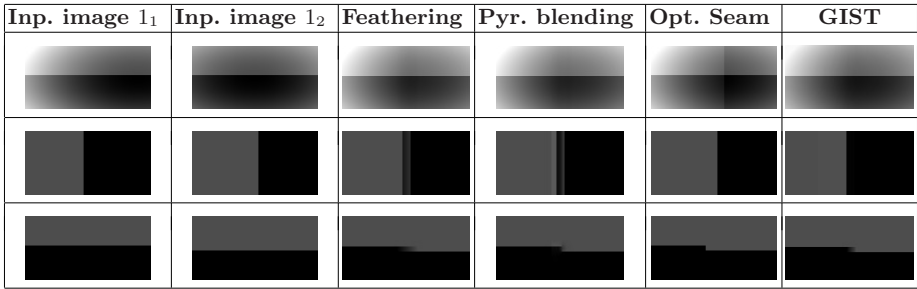


Fig. 2. Comparing stitching methods with various sources for inconsistencies between the input images. The left side of I_1 is stitched to right side of I_2 . Optimal seam methods produce a seam artifact in case of photometric inconsistencies between the images (first row). Feathering and pyramid blending produce double edges in case of horizontal misalignments (second row). In case there is a vertical misalignments (third row), the stitching is less visible with Feathering and GIST.

The stitching result I of GIST1 is defined as the minimum of E_p with respect to \hat{I} :

$$E_p(\hat{I}; I_1, I_2, W) = d_p(\nabla \hat{I}, \nabla I_1, \tau_1 \cup \omega, W) + d_p(\nabla \hat{I}, \nabla I_2, \tau_2 \cup \omega, U - W) \quad (1)$$

where U is a uniform image, and $d_p(J_1, J_2, \phi, W)$ is the distance between J_1, J_2 on ϕ :

$$d_p(J_1, J_2, \phi, W) = \sum_{\mathbf{q} \in \phi} W(\mathbf{q}) \| J_1(\mathbf{q}) - J_2(\mathbf{q}) \|_p^p \quad (2)$$

with $\| \cdot \|_p$ denoting the ℓ_p -norm.

The dissimilarity E_p between the images is defined by the distance between their derivatives. A dissimilarity in the gradient domain is invariant to the mean intensity of the image. In addition it is less sensitive to smooth global differences between the input images, e.g. due to non-uniformness in the camera photometric response and due to scene shading variations. On the overlap region ω , the cost function E_p penalizes for derivatives which are inconsistent with any of the input images. In image locations where both I_1 and I_2 have low gradients, E_p penalizes for high gradient values in the mosaic image. This property is useful in eliminating false stitching edges.

The choice of norm (parameter p) has implications on both the optimization algorithm and the mosaic image. The minimization of E_p (Equation 1) for $p \geq 1$ is convex, and hence efficient optimization algorithms can be used. Section 3 describes a minimization scheme for E_2 by existing algorithms, and a novel fast minimization scheme for E_1 . The mask image W was either a uniform mask (for E_1) or the Feathering mask (for E_2), which is linear with the signed-distance from the seam. The influence of the choice of p on the result image is addressed in the following sections, with the introduction of alternative stitching algorithms in the gradient domain.

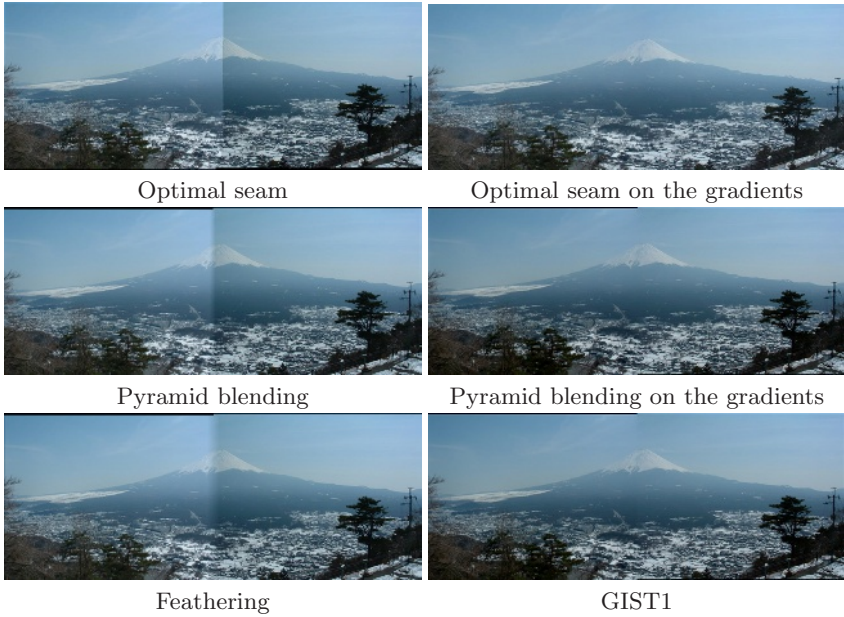


Fig. 3. Stitching in the gradient domain. The input images appear in Figure 1, with the overlap region marked by a black rectangle. With the image domain methods (top panels) the stitching is observable. Gradient-domain methods (bottom panels) overcome global inconsistencies.

2.2 GIST2: Stitching Derivative Images

A simpler approach is to stitch the derivatives of the input images:

1. Compute the derivatives of the input images $\frac{\partial I_1}{\partial x}, \frac{\partial I_1}{\partial y}, \frac{\partial I_2}{\partial x}, \frac{\partial I_2}{\partial y}$.
2. Stitch the derivative images to form a field $F = (F_x, F_y)$. F_x is obtained by stitching $\frac{\partial I_1}{\partial x}$ and $\frac{\partial I_2}{\partial x}$, and F_y is obtained by stitching $\frac{\partial I_1}{\partial y}$ and $\frac{\partial I_2}{\partial y}$.
3. Find the mosaic image whose gradients are closest to F . This is equivalent to minimizing $d_p(\nabla I, F, \pi, U)$ where π is the entire image area and U is a uniform image.

In stage (2) above, any stitching algorithm may be used. We have experimented with Feathering, pyramid blending [7], and optimal seam. For the optimal seam we used the algorithm in [2], finding the curve $x = f(y)$ that minimizes the sum of absolute differences in the input images. Stage (3), the optimization under ℓ_1, ℓ_2 , is described in Section 3. Unlike the GIST1 algorithm described in the previous section, we found minor differences in the result images when minimizing d_p under ℓ_1 and ℓ_2 .

2.3 Which Method to Use?

In the previous sections we presented several stitching methods. Since stitching results are tested visually, selecting the most appropriate method may be subject to personal taste. However, a formal analysis of properties of these methods is provided below. Based on those properties in conjunction with the experiments in Section 4, we recommend using GIST1 under ℓ_1 .

Theorem 1. *Let I_1, I_2 be two input images for a stitching algorithm, and assume there is a curve $x = f(y)$, such that for each $\mathbf{q} \in \{(f(y), y)\}$, $I_1(\mathbf{q}) = I_2(\mathbf{q})$. Let U be a uniform image. Then the optimal seam solution I , defined below, is a global minimum of $E_p(I; I_1, I_2, U)$ defined in Eq.1, for any $0 < p \leq 1$.*

$$I = \begin{cases} I_1(x, y) & x < f(y) \\ I_2(x, y) & x \geq f(y) \end{cases}$$

The reader is referred to [16] for a proof. The theorem implies that GIST1 under ℓ_1 is as good as the optimal seam methods when a perfect seam exists. Hence the power of GIST1 under ℓ_1 to overcome geometric misalignments similarly to the optimal seam methods. The advantage of GIST1 over optimal seam methods is when there is no perfect seam, for example due to photometric inconsistencies between the input images. This was validated in the experiments.

We also show an equivalence between GIST1 under ℓ_2 and Feathering of derivatives (GIST2) under ℓ_2 (Note that feathering derivatives is different from Feathering the images).

Theorem 2. *Let I_1, I_2 be two input images for a stitching algorithm, and let W be a Feathering mask. Let ω , the overlap region of I_1, I_2 , be the entire image (without loss of generality, as $W(\mathbf{q}) = 1$ for $\mathbf{q} \in \tau_1$, and $W = 0$ for $\mathbf{q} \in \tau_2$). Let I_{Gist} be the minimum of $E_2(I; I_1, I_2, W)$ defined in Eq. 1. Let F be the following field:*

$$F = W(\mathbf{q})\nabla I_1(\mathbf{q}) + (1 - W(\mathbf{q}))\nabla I_2(\mathbf{q})$$

Then I_{Gist} is the image with the closest gradient field to F under ℓ_2 .

The proof can be found in [16] as well. This provides insight into the difference between GIST1 under ℓ_1 and under ℓ_2 : Under ℓ_2 , the algorithm tends to mix the derivatives and hence blur the texture in the overlap region. Under ℓ_1 , the algorithm tends to behave similarly to the optimal seam methods, while reducing photometric inconsistencies.

3 Implementation Details

We have implemented a minimization for Equation 1 under ℓ_1 and under ℓ_2 .

Equation 1 defines a set of linear equations in the image intensities, with the derivative filters as the coefficients. Similarly to [12,13], we found that good results are obtained when the derivatives are approximated by forward-differencing

filters $\frac{1}{2} [1 \ -1]$. In the ℓ_1 case, the results were further enhanced by incorporating additional equations using derivative filters in multiple scales. In our experiments we added the filter corresponding to forward-differencing in the 2nd level of a Gaussian pyramid, obtained by convolving the filter $[1 \ 0 \ -1]$ with a vertical and a horizontal Gaussian filter ($\frac{1}{4} [1 \ 2 \ 1]$). Color images were handled by applying the algorithm to each of the color channels separately.

The minimum to Equation 1 under ℓ_2 with mask W is shown in [16] to be the image with the closest derivatives under ℓ_2 to F , the weighted combination of the derivatives of the input images:

$$F = \begin{cases} W(\mathbf{q})\nabla I_1(\mathbf{p}) & \mathbf{q} \in \tau_1 \\ W(\mathbf{q})\nabla I_1(x, y) + (1 - W(\mathbf{q}))\nabla I_2(x, y) & \mathbf{q} \in \omega \\ \nabla I_2(x, y) & \mathbf{q} \in \tau_2 \end{cases}$$

The solution can be obtained by various methods, e.g. de-convolution [12], FFT [17] or multigrid solvers [18]. The results presented in this paper were obtained by FFT.

As for the ℓ_1 optimization, we found using a uniform mask U to be sufficient. Solving the linear equations under ℓ_1 can be done by linear programming[19]:

$$\begin{aligned} &Min : \sum_i (z_i^+ + z_i^-) \\ &Subject \ to : Ax + (z^+ - z^-) = b, x \geq 0, z^+ \geq 0, z^- \geq 0 \end{aligned}$$

The entries in matrix A are defined by the coefficients of the derivative filters, and the vector b contains the derivatives of I_1, I_2 . x , is a vectorization of the result image.

The linear program was solved using LOQO[20]. A typical execution time for a 200×300 image on a Pentium 4 was around 2 minutes. Since no boundary conditions were used, the solution was determined up to a uniform intensity shift. This shift can be determined in various ways. We chose to set it according to the median of the values of the input image I_1 and the median of the corresponding region in the mosaic image.

3.1 Iterative ℓ_1 Optimization

A faster ℓ_1 optimization can be achieved by an iterative algorithm in the image domain. One way to perform this optimization is described in the following. Due to space limitation, we describe the algorithm when the forward differencing derivatives are used with kernel $\frac{1}{2} [1 \ -1]$. The generalization to other filters and a parallel implementation appear in [16]. Let Dx_j, Dy_j be the forward-differences of input image I_j . The optimization is performed as follows:

- Initialize the solution image I
- Iterate until convergence:
 - for all x,y in the image, update $I(x, y)$ to be:

$$2 * median(\cup_j \{ \begin{matrix} I(x + 1, y) - Dx_j(x, y), I(x - 1, y) + Dx_j(x - 1, y), \\ I(x, y + 1) - Dy_j(x, y), I(x, y - 1) + Dy_j(x, y - 1) \end{matrix} \}) \quad (3)$$

For an even number of samples, the median is taken to be the average of the two middle samples. In regions τ_j where a single image I_j is used, the median is taken on the predictions of $I(x, y)$ given its four neighbours and the derivatives of image I_j . For example, when the derivatives of image I_j are 0, the algorithm performs an iterated median filter of the neighbouring pixels. In the overlap region ω of I_1, I_2 , the median is taken over the predictions from both images.

At every iteration, the algorithm performs a coordinate descent and improves the cost function until convergence. As the cost function is bounded by zero, the algorithm always converges. However, although the cost function is convex, the algorithm does not always converge to the global optimum¹. To improve the algorithm convergence and speed, we combined it in a multi-resolution scheme using multigrid [18]. In extensive experiments with the multi-resolution extension the algorithm always converged to the global optimum.

4 Experiments

We have implemented various versions of GIST and applied them to panoramic mosaicing and object blending.

First, we compared GIST to existing image stitching techniques, which work on the image intensity domain: Feathering [6], Pyramid Blending [7], and 'optimal seam' (Implemented as in [2]). The experiments (Figure 3) validated the advantage in working in the gradient for overcoming photometric inconsistencies. Second, we compared the results of GIST1 (Section 2.1), GIST2 (Section 2.2) and the method by Perez. et. al. [10]. Results of these comparisons are shown, for example, in Figures 4,5, and analyzed in the following sections.

4.1 Stitching Panoramic Views

The natural application for image stitching is the construction of panoramic pictures from multiple input pictures. Geometrical misalignments between input images are caused by lens distortions, by the presence of moving objects, and by motion parallax. Photometric inconsistencies between input images may be caused by a varying gain, by lens vignetting, by illumination changes, etc.

The input images for our experiments were captured from different camera positions, and were aligned by a $2D$ parametric transformation. The aligned images contained local misalignments due to parallax, and photometric inconsistencies due to differences in illumination and in camera gain. Mosaicing results are shown in Figures 3,4,5. Figure 3 compares gradient methods vs. image domain methods. Figure 4,5 demonstrate the performance of the stitching algorithms when the input images are misaligned. In all our experiments GIST1 under ℓ_1 gave the best results, in some cases comparable with other methods: In Figure 4 comparable with Feathering, and in 5 comparable with 'optimal seam'.

¹ Consider an image whose left part is white and the right part is black. When applying the algorithm on the derivatives of this image, the uniform image is a stationary point.

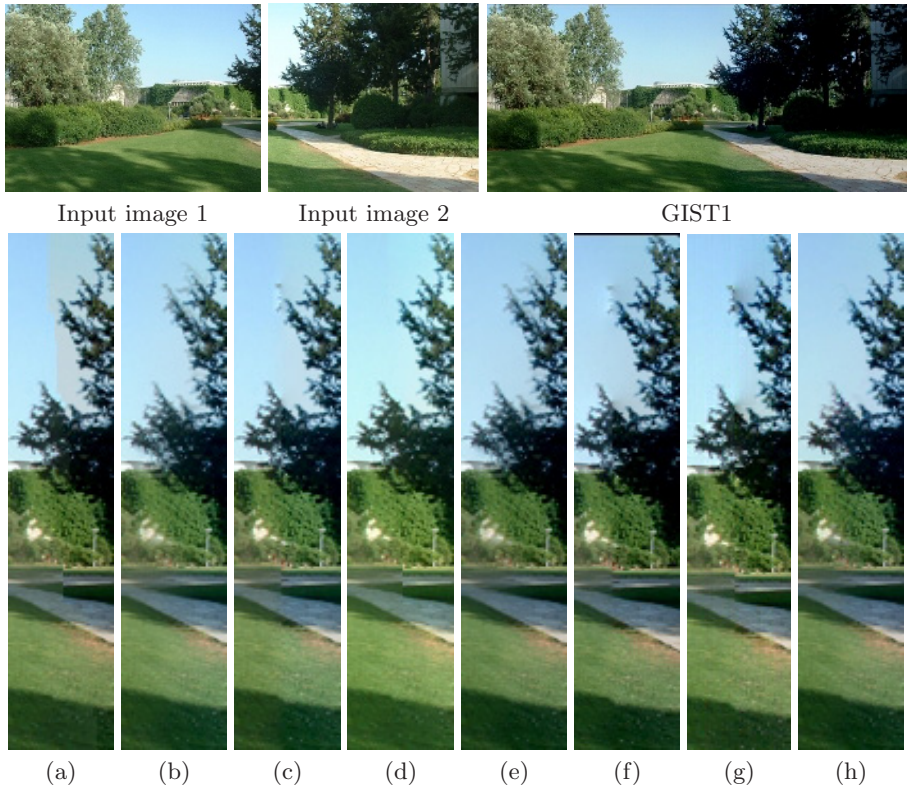


Fig. 4. Comparing various stitching methods. On top are the input image and the result of GIST1 under ℓ_1 . The images on bottom are cropped results of various methods. (a)-Optimal seam, (b)-Feathering, (c)-Pyramid blending, (d)-Optimal seam on the gradients, (e)-Feathering on the gradients, (f)-Pyramid blending on the gradients, (g)-Poisson editing [10] and (h) GIST1 - ℓ_1 . The seam is visible in (a),(c),(d),(g).

Whenever the input images were misaligned along the seam, GIST1 under ℓ_1 was superior to [10].

4.2 Stitching Object Parts

Here we combined images of objects of the same class having different appearances. Objects parts from different images were combined to generate the final image. This can be used, for example, by the police, in the construction of a suspect's composite portrait from parts of faces in the database. Figure 6 shows an example for this application, where GIST1 is compared to pyramid blending in the gradient domain. Another example for combination of parts is shown in Figure 7.



Fig. 5. A comparison between various image stitching methods. On top are the input image and the result of GIST1 under ℓ_1 . The images on bottom are cropped from the results of various methods. (a)-Optimal seam, (b)-Feathering, (c)-Pyramid blending, (d)-Optimal seam on the gradients, (e)-Feathering on the gradients, (f)-Pyramid blending on the gradients, (g)-Poisson editing [10] and (h) GIST1 - ℓ_1 . When there are large misalignments, optimal seam and GIST1 produce less artifacts.

5 Discussion

A novel approach to image stitching was presented, with two main components: First, images are combined in the gradient domain rather than in the intensity domain. This reduces global inconsistencies between the stitched parts due to illumination changes and changes in the camera photometric response. Second, the mosaic image is inferred by optimization over image gradients, thus reducing seam artifacts and edge duplications. Experiments comparing gradient domain stitching algorithms and existing image domain stitching show the benefit of stitching in the gradient domain. Even though each stitching algorithm works better for some images and worse for others, we found that GIST1 under ℓ_1 always worked well and we recommend it as the standard stitching algorithm. The use of the ℓ_1 norm was especially valuable in overcoming geometrical misalignments of the input images.

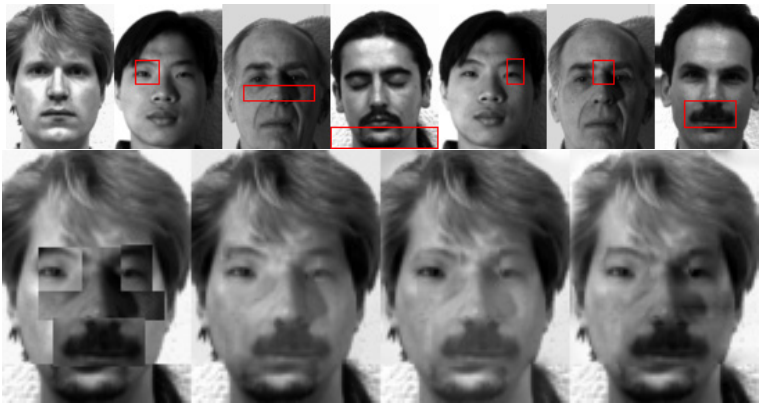


Fig. 6. A police application for generating composite portraits. The top panel shows the image parts used in the composition, taken from the Yale database. The bottom panel shows, from left to right, the results of pasting the original parts, GIST1 under ℓ_1 , GIST1 under ℓ_2 and pyramid blending in the gradient domain. Note the discontinuities in the eyebrows.

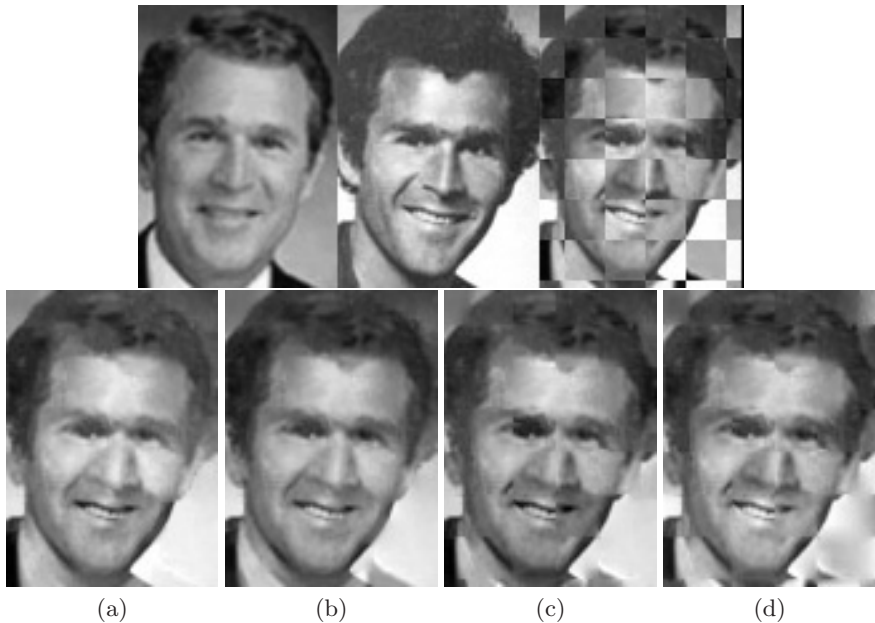


Fig. 7. A combination of images of George W. Bush taken at different ages. On top are the input images and the combination pattern. On the bottom left are, from left to right, the results of GIST1 Stitching under ℓ_1 (a) and under ℓ_2 (b), the results of pyramid blending in the gradient domain (c), and pyramid blending in the image domain(d).

The closest approach to ours was presented recently by Perez et. al. [10] for image editing. There are two main differences with this work: First, in this work we use the gradients of **both** images in the overlap **region**, while Perez et. al. use the gradients of the inserted object and the intensities of the background image. Second, the optimization is done under different norms, while Perez et. al. use the ℓ_2 norm. Both differences considerably influence the results, especially in misaligned textured regions. This is shown in Figures 5,4.

Image stitching was presented as a search for an optimal solution to an image quality criterion. The optimization of this criterion under norms ℓ_1, ℓ_2 is convex, having a single solution. Encouraged by the results obtained by this approach, we believe that it will be interesting to explore alternative criteria for image quality. One direction can use results on statistics of filter responses in natural images [21,22,23]. Another direction is to incorporate additional image features in the quality criterion, such as local curvature. Successful results in image inpainting[11,24] were obtained when image curvature was used in addition to image derivatives.

Acknowledgments. The authors would like to thank Dhruv Mahajan and Raanan Fattal for their help in the multigrid implementation, and Rick Szeliski for providing helpful comments.

References

1. Freeman, W., Pasztor, E., Carmichael, O.: Learning low-level vision. In: Int. Conf. on Computer Vision. (1999) 1182–1189
2. Efros, A., Freeman, W.: Image quilting for texture synthesis and transfer. Proceedings of SIGGRAPH 2001 (2001) 341–346
3. Milgram, D.: Computer methods for creating photomosaics. IEEE Trans. Computer **23** (1975) 1113–1119
4. Davis, J.: Mosaics of scenes with moving objects. In: CVPR. (1998) 354–360
5. Peleg, S., Rousso, B., Rav-Acha, A., Zomet, A.: Mosaicing on adaptive manifolds. IEEE Trans. on Pattern Analysis and Machine Intelligence **22** (2000) 1144–1154
6. Uyttendaele, M., Eden, A., Szeliski, R.: Eliminating ghosting and exposure artifacts in image mosaics. In: CVPR. (2001) II:509–516
7. Adelson, E.H., Anderson, C.H., Bergen, J.R., Burt, P.J., M., O.J.: Pyramid method in image processing. RCA Engineer **29(6)** (1984) 33–41
8. Peleg, S.: Elimination of seams from photomosaics. CGIP **16** (1981) 90–94
9. Fattal, R., Lischinski, D., Werman, M.: Gradient domain high dynamic range compression. Proceedings of SIGGRAPH 2001 (2002) 249–356
10. Perez, P., Gangnet, M., Blake, A.: Poisson image editing. SIGGRAPH (2003) 313–318
11. Ballester, C., Bertalmio, M., Caselles, V., Sapiro, G., Verdera, J.: Filling-in by joint interpolation of vector fields and gray levels. IEEE Trans. Image Processing **10** (2001)
12. Weiss, Y.: Deriving intrinsic images from image sequences. In: ICCV. (2001) II: 68–75

13. Tappen, M., Freeman, W., Adelson, E.: Recovering intrinsic images from a single image. In: NIPS. Volume 15., The MIT Press (2002)
14. Finlayson, G., Hordley, S., Drew, M.: Removing shadows from images. In: ECCV. (2002) IV:823
15. Levin, A., Zomet, A., Weiss, Y.: Learning to perceive transparency from the statistics of natural scenes. In: NIPS. Volume 15., The MIT Press (2002)
16. Levin, A., Zomet, A., Peleg, S., Weiss, Y.: Seamless image stitching in the gradient domain, hebrew university tr:2003-82, available on <http://leibniz.cs.huji.ac.il/tr/acc/2003/huji-cse-ltr-2003-82.blending.pdf> (2003)
17. Frankot, R., Chellappa, R.: A method for enforcing integrability in shape from shading algorithms. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **10** (1988) 439–451
18. Press, W., Flannery, B., Teukolsky, S., Vetterling, W.: *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge (UK) and New York (1992)
19. Chvátal, V.: *Linear Programming*. W.H. Freeman and CO., New York (1983)
20. Vanderbei, R.: *Loqo*, <http://www.princeton.edu/rvdb/> (2000)
21. Mallat, S.: A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **11** (1989) 674–693
22. Simoncelli, E.: Bayesian denoising of visual images in the wavelet domain. *BIWBM* **18** (1999) 291–308
23. Wainwright, M., Simoncelli, E., Willsky, A.: Random cascades of gaussian scale mixtures for natural images. In: *Int. Conf. on Image Processing*. (2000) I:260–263
24. Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C.: Image inpainting. In: *SIG-GRAPH*. (2000)