Adobe Technical Report, March 2009

A Perceptually Motivated Method to Control Reconstruction Errors in Gradient-based Image Compositing

Sylvain Paris

Abstract

We propose a simple algorithm to control the spatial location of reconstruction errors inherent in gradient-based image compositing. We build upon the classical Poisson equation and add a weighting term that controls where reconstruction errors can occur. We define this term in such a way that residuals are mainly located in textured regions where they are less visible. Our approach is independent of how the composited gradient field has been built and is complementary to the methods that focus on this aspect. Our approach retains the simplicity of the traditional Poisson equation while producing more pleasing composites.

1. Introduction

Gradient compositing is a popular technique for stitching images together, either for local edits [2, 10, 12, 17] or panorama stitching [1,14]. These methods first delineate the composited regions, then compute a target gradient field and boundary conditions from these regions, and finally solve the Poisson equation to reconstruct an image. A major issue with gradient-based compositing is that the combined gradient field may not be integrable, that is, there may not exist an image which gradients match the target field and the specified boundary conditions. Existing work mitigates this aspect by carefully selecting and combining the merged regions. However, when the combined images are widely different, this strategy may not be sufficient. Generated target gradient fields may be far from integrable, yielding color leaks and halos typical of Poisson-based methods. In this paper, we focus on these artifacts and propose a method to control their location so that they are less conspicuous. In other words, errors are still present our results, but they are less objectionable than the errors generated by the standard method.

Our approach also aims for numerical simplicity because the available computational budget to reconstruct the image from the gradients is often limited. For instance, panorama stitching generates problems on the order of several tens of megapixels [1] and the Photoshop Healing Brush runs at interactive rates [10]. Our method relies on a sparse linear system that can be solved without introducing a significant overhead compared to the standard technique.

To summarize, our approach is best suited for interactive tools and for miscalibrated panoramas where errors are often large and inevitable and for which performance and scalability are crucial issues. In these scenarios, our method yields more satisfying outputs than the classical Poisson equation without sacrificing computational efficiency.

1.1. Related Work

Our approach is complementary and orthogonal to the methods that determine the target gradient field by determining the regions boundaries and combining the data from both regions [2, 10, 12, 14, 17]. We focus on reconstructing the final image from the target gradient field and specified boundaries. The most often used option is to seek for an image I that approximates the target field $\mathbf{v}$ in a least-squares sense (with ∇ the gradient operator):

$$\operatorname{argmin}_I \int \|\nabla I - \mathbf{v}\|^2 \tag{1}$$

which can be minimized by solving the Poisson equation:

$$\Delta I - \operatorname{div}(\mathbf{v}) = 0 \tag{2}$$

where Δ is the Laplacian operator $\partial^2/\partial x^2 + \partial^2/\partial y^2$ and *div* is the divergence operator $\partial/\partial x + \partial/\partial y$. To solve this equation, one also needs boundary conditions that depend on the application. For instance, for cutting-and-pasting, the values of the boundary pixels provide these conditions. The equation is named after Poisson who introduced it in Physics to describe potentials in uniform regions [16]. Our contribution is to adapt it to nonuniform image content.

Variants of the Poisson Equation Our work shares similarities with the method of Lalonde *et al*. [13] who propose to take the image gradient magnitude into account during the reconstruction process. However, Lalonde's scheme is nonlinear and is presented as a heuristic in the context of a larger problem; it is neither studied nor validated independently of the boundary selection process. Our work fills in this gap and proposes a linear variant that can be solved with any linear solver. Our work is also related to the unpublished work by Bhat *et al*. [6] who follow an approach related to ours but use it for image stylization whereas we study compositing artifacts and rely on visual masking to minimize them.

Shape from Shading Beside image compositing, the Poisson equation has been largely used and adapted to solve the shape-from-shading problem. We refer to the recent work of Agrawal et al. [3] and Harker and O'Leary [11] and the references therein for detail. In particular, the form of our equations are similar to the anisotropic scheme proposed by Agrawal. Yet, there are fundamental differences between images and depth maps. Most importantly for our work, the L_2 norm classically used on depth maps poorly reflects the human perception of images [5,7]. Shape-from-shading techniques applied directly to images would not account for our perception of images. Our contribution is to propose a scheme relying on domain-specific knowledge on image compositing and human perception.

Perceptual Rendering Our strategy is inspired by recent work on perceptual rendering that simplifies 3D scenes in ways that have a minor impact on the viewer's perception [8, 18–20] (and references therein). Our approach follows a similar strategy but applied to images. The fundamental difference between perceptual rendering and our work is that these rendering techniques can produce error-free images given sufficient time – introducing errors in non salient regions speeds up the computation but could be avoided. In comparison, we are given a target gradient field that cannot be integrated, that is, we must introduce errors in order to reconstruct the final image. Yet, as perceptual rendering, we put these errors in non salient regions.

1.2. Contributions

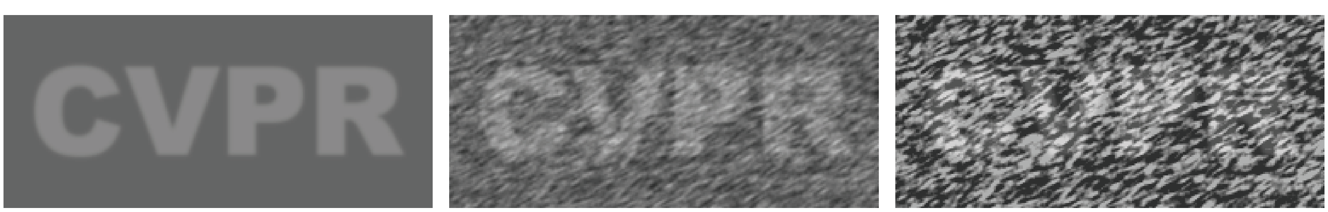
Compared to existing work, our main contribution is a method driven by perceptual criteria to control the location of integration residuals in gradient-based compositing. We argue that locating residuals in textured areas reduces their visual impact because it leverages visual masking and respects the image structure. We demonstrate our approach with a weighted version of the Poisson equation. Our algorithm is computationally lightweight so that it is widely usable. Furthermore, it achieves visually pleasing results.

2. Controlling the Location of the Residuals

Given a target gradient field $\mathbf{v}$ with boundary conditions, we aim for producing an image with a gradient field ∇I as close as possible to $\mathbf{v}$. We are interested in the case where we cannot exactly fulfill our objective and must introduce errors. Our strategy is to locate the residuals as much as possible in regions where they will be the least objectionable. Intuitively, we want to avoid errors in smooth regions such as the sky where they produce color leaks and halos, and put them in textured areas where visual masking will "hide" them. Figures 1 and 2 illustrate this effect.

(a) dots on a gray background (b) same dots on a photograph

Figure 2. Visual masking on a photograph. We show the same dots on a uniform background (a) and on a photograph (b). Only the dots on the sky are visible in the photo, the two dots on the foliage are not visible although they are identical to the ones shown on the left. The color of the blue and green dots have been picked from the sky and the foliage to make the configuration symmetric with respect to color. We informally showed the right image to a few people ignoring how it had been edited. No one saw the dots in the foliage, even after close observation.



(a) no texture (b) low contrast texture (c) high contrast texture

Figure 1. Example of visual masking. When the CVPR characters are overlaid with texture, they become difficult to read although the contrast of the characters themselves is the same. With the most contrasted texture on the right, the characters are almost invisible. Our approach leverages this perceptual effect to hide integration residuals in textured areas where they are the least visible.

2.1. Adapting the Poisson Equation

Let's assume that we have a scalar map M with high values in regions where errors would be visible and low values otherwise. We discuss later how to compute such a function. Given M, we modulate the least-squares strength so that we penalize less the regions where we prefer the residuals to be, that is, regions with low M values:

$$\operatorname{argmin}_I \int M \, \|\nabla I - \mathbf{v}\|^2 \tag{3}$$

Since we want to reduce the difference between ∇I and $\mathbf{v}$, M has to be strictly positive everywhere.

A Linear Solution To obtain a formula similar to the Poisson equation (2), we apply the Euler-Lagrange formula [4]. Recall that we aim at a simple solution. Thus, we make the choice of using a map M that is not a function of I nor ∇I so that we obtain the following linear system:

$$\operatorname{div}\big(M \, (\nabla I - \mathbf{v})\big) = 0 \tag{4}$$

In the following section, we show that, although this equation is simple, it has good properties. Later, we demonstrate that it produces visually pleasing results. We also discuss nonlinear alternatives in Section 3.

2.2. Analysis of the Residual Structure

Independently of the actual definition of M, we can already show that the residuals produced by our approach has a structure aligned with the image content. Wang *et al*. [21] have demonstrated that such structural similarity produces more acceptable results.

Influence of the Location Map To better understand the role of M, we distribute the divergence in Equation 4:

$$M \operatorname{div}(\nabla I - \mathbf{v}) + \nabla M \cdot (\nabla I - \mathbf{v}) = 0 \tag{5}$$

With $M \neq 0$, the relation: $\operatorname{div}(\nabla I) = \Delta I$, and the logarithmic gradient: $\nabla M / M = \nabla \log M$, we obtain:

$$\underbrace{\Delta I - \operatorname{div}(\mathbf{v})}_{\text{Poisson term}} \;+\; \underbrace{\nabla \log M \cdot (\nabla I - \mathbf{v})}_{\text{new term}} \;=\; 0 \tag{6}$$

The left term is the same as the standard Poisson equation (2) while the right term is new. Remarkably, this new term is driven by $\nabla \log M$, we build the following study on this point.

In regions where M is constant, the right term is null because $\nabla \log M = 0$. Our scheme behaves similarly to the Poisson equation: it uniformly spreads the residuals. In the other regions where $\nabla \log M \cdot (\nabla I - \mathbf{v}) \neq 0$, our results do not satisfy the Poisson equation (2), that is $\Delta I - \operatorname{div}(\mathbf{v}) \neq 0$. It allows for larger variations in the residual field. This can occur only where $\nabla \log M \neq 0$, that is large residual variations are aligned with edges of M. Importantly, this relationship is not reciprocal and M edges do not necessarily generate residual edges since the dot product can be null even if $\nabla \log M \neq 0$. For instance, this prevents spurious residual edges in regions where I has the desired gradient $\mathbf{v}$, that is, $\nabla I - \mathbf{v} = 0$.

Structure of the Integration Residuals We have shown that the residuals are smooth where M is smooth and can vary only at M edges. Furthermore, we assume that M has the same structure as the image, that is, M edges are aligned with image contours. This hypothesis holds in practice because we compute M from $\mathbf{v}$ that is close to ∇I. Combining these arguments, we can conclude that our method creates a residual field with a structure similar to the image which is known to be less objectionable [21]. This result can be seen in Figure 3. Our method produces almost constant residuals in the sky and on the foliage where the amount of texture does not vary. Changes mostly happen at the silhouette of the tree and are not noticeable. In comparison, the Poisson residual varies everywhere. Although the variations are smooth, they cause visible leaks and halos in the sky because of the absence of texture.

2.3. Computing the Location Map

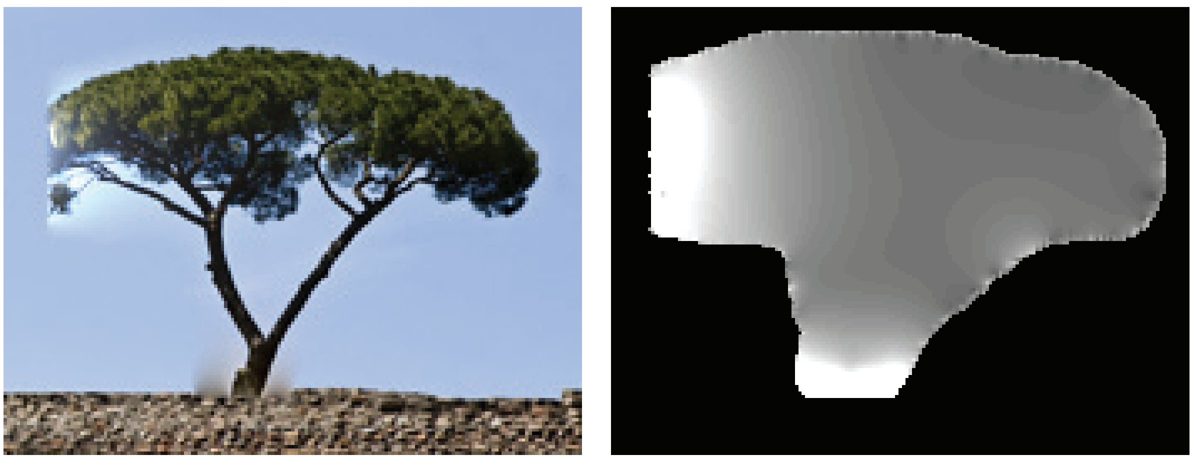
In this section, we explain how to actually compute the scalar map M. We provide a function that fulfills our objectives and keep the exploration of the space of all possible options as future work.

As previously discussed, M should be positive and have high values where artifacts would be the most visible and low values where they would less conspicuous. To preserve the linearity of the solution, M should not be a function of I nor ∇I. We also impose $M \leq 1$ so that the convergence properties of the system (4) is similar to the Poisson equation (2). Finally, we seek a computationally inexpensive solution so that our approach can be widely used.

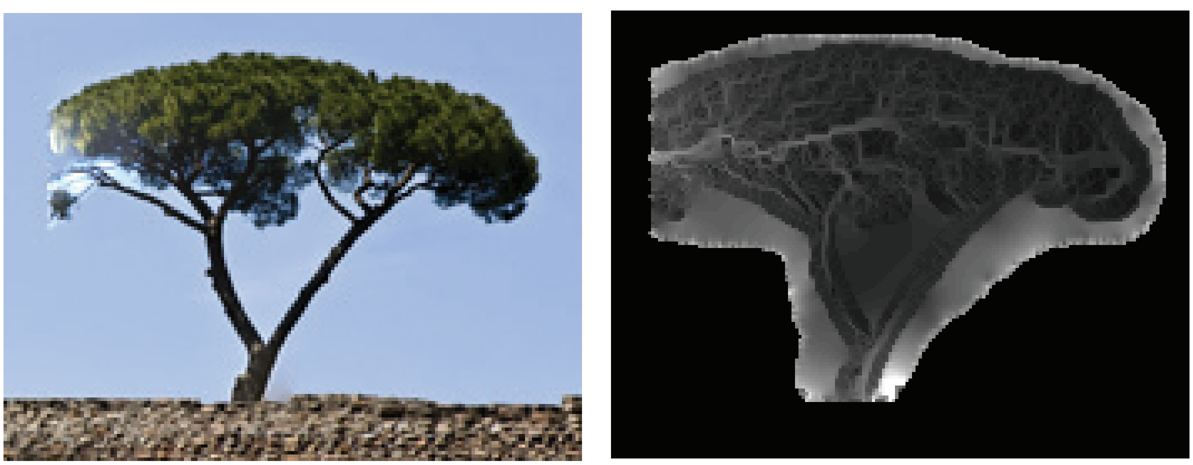
Our Model We use a simple model of visual masking based on image variations. We consider highly varying areas to be regions where the errors will have a lower visual impact, thus we assign them lower M values. Since we cannot use I nor ∇I, we use the target gradient field $\mathbf{v}$ to estimate the image variations. Although $\mathbf{v}$ described the target image variations, not the actual variations, it provides a sufficiently good estimate while keeping the equations linear. We propose the following expression for M:

$$M = 1 - p \; \min\left(1, \frac{\|\mathbf{v}\| \, n(\|\mathbf{v}\|)}{G_\sigma \otimes \|\mathbf{v}\|}\right) \tag{7}$$

(a) input (the tree has been pasted) (b) scalar map M (white = 1 and black = 0)



(c) standard Poisson reconstruction and integration residual (higher is brighter)



(d) our reconstruction and integration residual (higher is brighter)

Figure 3. Pasted tree. The left side of the tree has been erroneously cut and the trunk has been pasted within the ground region. The standard Poisson reconstruction introduces halos near the foliage and leaks near the trunk. Although these artifacts are not removed by our approach, they are significantly reduced. In particular, the outer halo around the foliage has disappeared and the leak near the trunk is mostly gone. Our approach relies on the scalar map M shown on the top right (cf. text for detail). Compared to a standard Poisson equation that uniformly spreads the integration errors (middle right), our approach accounts to the image content (bottom right).

where $G_\sigma \otimes \|v\|$ is the convolution of the amplitudes of $\mathbf{v}$ with a Gaussian kernel and represents the average local variations. In practice, we found that small neighborhoods performed well and we use $\sigma = 8$ in all our examples. Locations with a ratio $\|\mathbf{v}\| \,/\, G_\sigma \otimes \|\mathbf{v}\| \geq 1$ have variations above their local average and we consider them suitable locations where to put integration residuals. We treat them equally as indicated by the *min*. However, this method is sensitive to noise in smooth regions because $G_\sigma \otimes \|\mathbf{v}\|$ is small. We address this issue with the function n that is equal to 1 except for small values of $\|\mathbf{v}\|$ for which $n = 0$. In practice, n is a smooth-step function such that $n = 0$ for $\|\mathbf{v}\|$ less than 2% of the intensity scale, $n = 1$ above 4%, and a smooth variation in between. The variable p is a global parameter that indicates how much we control the residual location. For instance, $p = 0$ corresponds to no control, that is, to the standard Poisson equation, whereas larger values impose more control. p has to be strictly smaller than 1 to keep $M > 0$. We found that values close to 1 performs better in practice. We use $p = 0.999$ in all our examples. Finally, the $1 - \ldots$ ensures that M has low values in smooth regions and high values in textured regions while remaining between 0 and 1.

Our experiments show that this function yields satisfying results. Compared to the Poisson equation (2), the Gaussian convolution by a small kernel is the only significant additional cost, which is an inexpensive operation. Nonetheless, if one has a larger computational budget, nothing prevents our approach to use more sophisticated models inspired for instance by the work of Daly [7] and Aydin *et al*. [5]

3. Relationship with Existing Methods

For this section, we make explicit the variables used to define M, that is, Equation 3 becomes: $\int M(\mathbf{v})\, \|\nabla I - \mathbf{v}\|^2$, and Equation 4: $\mathrm{div}\big(M(\mathbf{v})\,(\nabla I - \mathbf{v})\big) = 0$. We discuss the relationships between our work and related methods independently of the actual definition of M.

The Poisson Equation and its Variants Rewriting the Poisson equation (2) as: $\mathrm{div}(\nabla I - \mathbf{v}) = 0$, we see that our linear system has the same complexity since we do not introduce new unknowns nor new coefficients in the system, we only reweight the coefficients. With our notation, Lalonde's reconstruction method [13] minimizes $\int M(\nabla I)\, \|\nabla I - \mathbf{v}\|^2$ plus an attachment term. This functional takes the variations of the actual image into account but it sacrifices the least-squares formulation for a more complex, nonlinear one. Agrawal *et al*. [3] propose another nonlinear alternative where the least-squares norm is replaced by a robust norm. This formulation requires an iterative solver and would be nontrivial to adapt to large-scale problems [1]. Agrawal *et al*. also describe an anisotropic variant that is linear. However, this work focuses on shape-from-shading while we address issues specific to image compositing. In particular, the uniform norms used on depth maps poorly model human perception. Our work specifically addresses this issue by locally estimating the visual impact of integration residuals on the image quality.

Edge-preserving Filtering Our method is also related to Farbman's edge-preserving filter [9] that minimizes an attachment term plus $\int M(I_0)\, \|\nabla I\|^2$ where I_0 is the input image. Note that Farbman projects the formula on the x and y axes but we believe that it does not have a major impact on the results. More importantly, Farbman's method and ours share the same idea of using a modulation M that depends on fixed quantities that preserve the least-squares nature of the problem; Farbman uses the input image I_0 and we use the target gradient field $\mathbf{v}$. Finally, our work has common points with Perona and Malik's nonlinear anisotropic diffusion filter [15]: $\partial I/\partial t = \mathrm{div}\big(M(\nabla I)\, \nabla I\big)$. The difference is that our modulation term M is not a function of the image I which makes our equation linear, and we have a term $\nabla I - \mathbf{v}$ instead of ∇I, which can be interpreted as Perona and Malik "diffuse gradients" whereas we "diffuse integration residuals".

4. Results

We demonstrate our approach on a two typical scenarios: hand-made compositing (Figures 3 and 6) and misaligned panoramas (Figure 5). To better demonstrate our method, we created large errors on purpose. Although in some cases, automatic boundary refinement would be able to better adjust the boundary [2, 12], this is not always possible. For instance in Figure 6, pasting the seagull behind the street light would challenge most algorithms. Also, as previously discussed, reconstruction residuals remain in our outputs, that is, the produced images still contain limited halos and leaks. Nonetheless, the visual impact of those is greatly reduced by our method. As an example, the leaks near the trunk of the pasted tree are mostly gone and the halo around its foliage is confined in a much smaller region (Figure 3).

Limitation If the image has no texture near the compositing mismatches, there is nowhere to "hide" the integration residuals. In that case, our method does not improve the composite (Figure 4).

Conclusion We have exposed a method to control the location of integration residuals inherent in gradient-based methods. Our approach reduces the visual impact of these residuals. In addition, it is simple and can be easily added to existing software.

(a) Poisson reconstruction (b) our reconstruction

Figure 4. Pasted seagull. In absence of texture near compositing mismatches, our method is not able to produce better results than a standard Poisson reconstruction.

Acknowledgment The author thanks Kavita Bala for her inspiring talk on perception and rendering, the Adobe CTL members for their feedback, and Todor Georgiev for his analysis that appears in Section 2.2.

References

[1] A. Agarwala. Efficient gradient-domain compositing using quadtrees. *ACM Transactions on Graphics*, 26(3), 2007. Proceedings of the ACM SIGGRAPH conference.

[2] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. H. Salesin, and M. F. Cohen. Interactive digital photomontage. *ACM Transactions on Graphics*, 23(3):294–302, July 2004. Proceedings of the ACM SIGGRAPH conference.

[3] A. Agrawal, R. Raskar, and R. Chellappa. What is the range of surface reconstructions from a gradient field? In *Proceedings of the European Conference on Computer Vision*, 2006.

[4] G. Aubert and P. Kornprobst. *Mathematical problems in image processing: Partial Differential Equations and the Calculus of Variations*, volume 147 of *Applied Mathematical Sciences*. Springer, 2002.

[5] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H.-P. Seidel. Dynamic range independent image quality assessment. *ACM Transactions on Graphics*, 27(3), 2008. Proceedings of the ACM SIGGRAPH conference.

[6] P. Bhat, C. L. Zitnick, M. Cohen, and B. Curless. Gradientshop: A perceptually-motivated optimization-framework for image and video processing. Technical report, University of Washington, 2008.

[7] S. Daly. *Digital images and human vision*, chapter The visible differences predictor: an algorithm for the assessment of image fidelity. MIT Press, 1993.

[8] G. Drettakis, N. Bonneel, C. Dachsbacher, S. Lefebvre, M. Schwarz, and I. Viaud-Delmon. An interactive perceptual rendering pipeline using contrast and spatial masking. *Rendering Techniques*, 2007.

[9] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Transactions on Graphics*, 27(3), 2008. Proceedings of the ACM SIGGRAPH conference.

[10] T. Georgiev. Covariant derivatives and vision. In *Proceedings of the European Conference on Computer Vision*, 2006.

[11] M. Harker and P. O'Leary. Least squares surface reconstruction from measured gradient fields. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. IEEE, 2008.

[12] J. Jia, J. Sun, C.-K. Tang, and H.-Y. Shum. Drag-and-drop pasting. *ACM Transactions on Graphics*, 25(3), July 2006. Proceedings of the ACM SIGGRAPH conference.

[13] J.-F. Lalonde, D. Hoiem, A. Efros, C. Rother, J. Winn, and A. Criminisi. Photo clip art. *ACM Transactions on Graphics*, 26(3), July 2007. Proceedings of the ACM SIGGRAPH conference.

[14] A. Levin, A. Zomet, S. Peleg, and Y. Weiss. Seamless image stitching in the gradient domain. In *Proceedings of the European Conference on Computer Vision*, 2006.

[15] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions Pattern Analysis Machine Intelligence*, 12(7):629–639, July 1990.

[16] S. D. Poisson. in *Bulletin de la société philomatique*, 1813.

[17] P. Pérez, M. Gangnet, and A. Blake. Poisson image editing. *ACM Transactions on Graphics*, 22(3), July 2003. Proceedings of the ACM SIGGRAPH conference.

[18] G. Ramanarayanan, K. Bala, and J. Ferwerda. Perception of complex aggregates. *ACM Transactions on Graphics*, 27(3), 2008. Proceedings of the ACM SIGGRAPH conference.

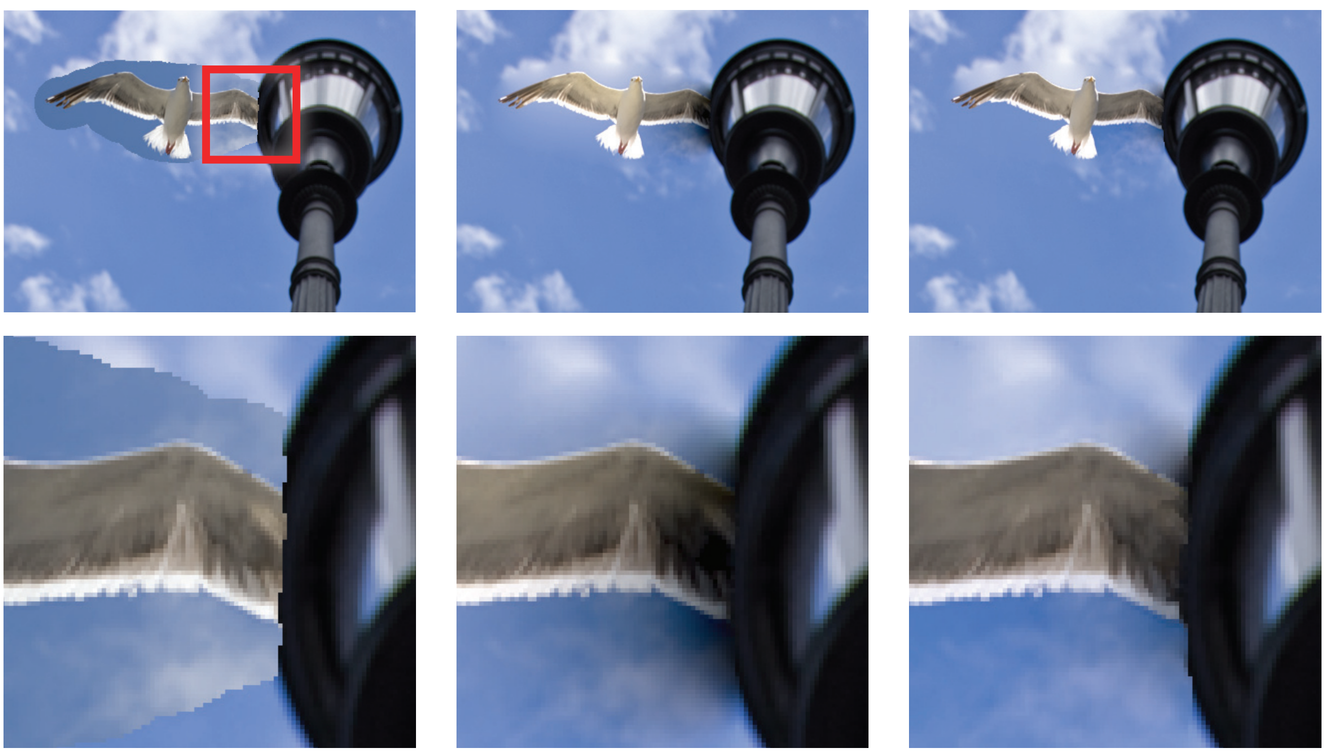
[19] G. Ramanarayanan, J. Ferwerda, B. Walter, and K. Bala. Visual equivalence: Towards a new standard for image fidelity. *ACM Transactions on Graphics*, 26(3), 2007. Proceedings of the ACM SIGGRAPH conference.

[20] P. Vangorp, J. Laurijssen, and P. Dutré. The influence of shape on the perception of material reflectance. *ACM Transactions on Graphics*, 26(3), 2007. Proceedings of the ACM SIGGRAPH conference.

[21] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 2004.

(a) input (misaligned panorama) (b) standard Poisson reconstruction (c) our reconstruction

Figure 5. Misaligned panorama. With a standard Poisson reconstruction, the error on the roof leaks in the sky and the transition on the riverbank remains hard. Our approach removes the leaks and generates a softer, less objectionable transition.



(a) input (the seagull has been pasted) (b) standard Poisson reconstruction (c) our reconstruction

Figure 6. Pasted seagull. The boundary between the right wing of the seagull and the light is not accurate. A standard Poisson reconstruction generates leaks in the sky whereas our method produces an acceptable result.