Perception-based energy functions in seam-cutting

Nan Li, Tianli Liao, and Chao Wang

Abstract—Image stitching is challenging in consumer-level photography, due to alignment difficulties in unconstrained shooting environment. Recent studies show that seam-cutting approaches can effectively relieve artifacts generated by local misalignment. Normally, seam-cutting is described in terms of energy minimization, however, few of existing methods consider human perception in their energy functions, which sometimes causes that a seam with minimum energy is not most invisible in the overlapping region. In this paper, we propose a novel perception-based energy function in the seam-cutting framework, which considers the nonlinearity and the nonuniformity of human perception in energy minimization. Our perception-based approach adopts a sigmoid metric to characterize the perception of color discrimination, and a saliency weight to simulate that human eyes incline to pay more attention to salient objects. In addition, our seam-cutting composition can be easily implemented into other stitching pipelines. Experiments show that our method outperforms the seam-cutting method of the normal energy function, and a user study demonstrates that our composed results are more consistent with human perception.

Index Terms—Image stitching, seam-cutting, energy function, human perception.

I. INTRODUCTION

Image stitching is a well studied topic in computer vision [1], which mainly consists of alignment [2–5], composition [6–10] and blending [11–13]. In consumer-level photography, it is difficult to achieve perfect alignment due to unconstrained shooting environment, so image composition becomes the most crucial step to produce artifacts-free results.

Seam-cutting [14–18] is a powerful composition method, which intends to find an invisible seam in the overlapping region of aligned images. Mainstream algorithms usually express the problem in terms of energy minimization and minimize it via graph-cut optimization [19–21]. Normally, for a given overlapping region of aligned images, different energy functions correspond to different seams, and certainly correspond to different composed results (see Fig. 1). Conversely, in order to obtain a plausible stitching result, we desire to define a perception-consistent energy function, such that the most invisible seam possesses the minimum energy.

Recently, many efforts have been devoted to seam-cutting by penalizing the photometric difference using various energy functions. A Euclidean-metric color difference is used in [14] to define the smoothness term in their energy function, and a gradient difference is taken into account in [15]. Eden et al. [16] proposed an energy function that allows for large motions and exposure differences, but the camera setting is required. Jia and Tang [17] associated the smoothness term with gradient smoothness and gradient similarity, to reduce structure complexity along the seam. Zhang et al. [18] combined alignment errors and a Gaussian-metric color difference in their energy function, to handle misaligned areas with similar colors. However, few of existing methods consider human perception in their energy functions, which sometimes causes that a seam with minimum energy is not most invisible in the overlapping region.

Seam-cutting has also been applied in image alignment. Gao et al. [22] proposed a seam-driven image stitching framework, which finds a best homography warp from some candidates with minimal seam costs instead of minimal alignment errors. Zhang and Liu [23] combined homography and content-preserving warps to locally align images, where seam costs are used as a quality metric to predict how well a homography enables plausible stitching. Lin et al. [24] proposed a seam-guided local alignment, which iteratively improves warping by adaptive feature weighting according to their distances to current seams.

In this paper, we propose a novel seam-cutting method via a perception-based energy function, which takes the nonlinearity and the nonuniformity of human perception into account. Our proposed method consists of three stages (see Fig. 2). In the first stage, we calculate a sigmoid-metric color difference of the given overlapping region as the smoothness term, to characterize the perception of color discrimination. Then, we calculate an average pixel saliency of the given overlapping region as the saliency weight, to simulate that human eyes incline to pay more attention to salient objects. Finally, we minimize the perception-based energy function by the graph-cut optimization, to obtain the seam and the corresponding composed result. Experiments show that our method outper-
Major contributions of the paper are summarized as follows.

1) We proposed a novel perception-based energy function in the seam-cutting framework.
2) Our composition method can be easily implemented into other stitching pipelines.

II. APPROACH

In this section, we first show more details of the normal seam-cutting framework, then a novel perception-based energy function is described, and finally we propose our seam-cutting framework.

A. Normal Seam-cutting Framework

Given a pair of aligned images denoted by $I_0$ and $I_1$, let $\mathcal{P}$ be their overlapping region and $\mathcal{L} = \{0, 1\}$ be a label set, where “0” corresponds to $I_0$ and “1” corresponds to $I_1$, then a seam means assigning a label $l_p \in \mathcal{L}$ to each pixel $p \in \mathcal{P}$. The goal of seam-cutting is to find a labeling $l$ (i.e., a map from $\mathcal{P}$ to $\mathcal{L}$) that minimizes the energy function

$$E(l) = \sum_{p \in \mathcal{P}} D_p(l_p) + \sum_{(p,q) \in \mathcal{N}} S_{p,q}(l_p, l_q),$$

where $\mathcal{N} \subset \mathcal{P} \times \mathcal{P}$ is a neighborhood system of pixels. The data term $D_p(l_p)$ represents the cost of assigning a label $l_p$ to a pixel $p \in \mathcal{P}$, and the smoothness term $S_{p,q}(l_p, l_q)$ represents the cost of assigning a pair of labels $(l_p, l_q)$ to a pair of pixels $(p, q) \in \mathcal{N}$.

The data term is defined as

$$D_p(l_p) = \begin{cases} 0, & \text{if } p \in \partial I_k \cap \partial \mathcal{P}, \quad \text{if } p \in \partial I_0 \cap \partial \mathcal{P}, \\ \mu, & \text{otherwise}, \end{cases}$$

where $\mu$ is a very large penalty to avoid mislabeling, $\partial I_k \cap \partial \mathcal{P}$ is the common border of $I_k$ ($k = 0, 1$) and $\mathcal{P}$ (marked in red and blue respectively in Fig. 1(a)). In fact, the data term $D_p(l_p)$ fixes the endpoints of the seam as the intersections of the two colored polylines.

The smoothness term is defined as

$$S_{p,q}(l_p, l_q) = \frac{1}{2} |l_p - l_q| (I_\ast(p) + I_\ast(q)),$$

where $I_\ast(\cdot)$ denotes the Euclidean-metric color difference (see Fig. 2(b)).

Finally, the normal energy function (1) is minimized by graph-cut optimization [19] to obtain the seam (see Fig. 2(e)) and the composed result (see Fig. 2(g)). Obviously, the defi-
nition of the energy function plays the most important role in the
seam-cutting framework.

B. Perception-based Energy Function

In experiments, the seam denoted by \( l_s \), that minimizes the
normal energy function \((1)\) is sometimes not most invisible in \( \mathcal{P} \).
In other words, there exists a seam denoted by \( l_t \), that is
more invisible but has a greater energy than \( l_s \) (see Fig. 2(c) and (f)).
Therefore, we desire to define a perception-consistent
energy function, such that the most invisible seam possesses
the minimum energy.

1) Sigmoid metric: Fig. 3 shows a toy example where \( l_s \) is
not most invisible. In fact, the seam \( l_s \) shown in (b) crosses the
local misalignment area (marked in light blue in (a)), because
the Euclidean-metric color difference does not give it a large
enough penalty. In contrast, the seam \( l_t \) shown in (d) avoid
the local misalignment area (marked in red in (c)), because
the sigmoid-metric color difference successfully distinguish it
from the alignment area.

In particular, the perception of colors is nonlinear as it has
a color discrimination threshold, which means human eyes
cannot differentiate some colors from others even if they are
different. Let \( \tau \) denote the threshold, the perception of color
discrimination can be characterized as

- if \( I_s(\cdot) < \tau \), color difference is invisible,
- if \( I_s(\cdot) \approx \tau \), sensitivity of discrimination rises rapidly,
- if \( I_s(\cdot) > \tau \), color difference is visible.

We want to define a quality metric to measure the visibility
of color difference, such that the cost of invisible terms
approximates zero while the cost of visible terms approximates one. Fortunately, the sigmoid function

\[
sigmoid(x) = \frac{1}{1 + e^{-4\kappa(x-\tau)}},
\]

is a suitable quality metric for our purpose.

Now, the smoothness term is modified as

\[
\tilde{S}_{p,q}(l_p, l_q) = \frac{1}{2} \| l_p - l_q \| (I_t(p) + I_t(q)),
\]

where \( I_t(\cdot) \) denotes the sigmoid-metric color difference.

\[
I_t(\cdot) = \text{sigmoid}(I_s(\cdot)),
\]

2) Saliency weights: Fig. 4 shows another toy example
where \( l_s \) is not most invisible. In fact, seams \( l_s \) and \( l_t \) shown in
(b) and (d) respectively, both cross the local misalignment area.
Though the energy of \( l_t \) is greater, it is more invisible than \( l_s \) in aspect of human perception, because the location
where its artifact arises is less remarkable than \( l_s \).

In particular, the perception of images is nonuniform, which
means that human eyes incline to pay more attention to salient
objects. Thus artifacts in salient regions are more remarkable
than artifacts in non-salient regions. In order to benefit from
these observations, we define a saliency weight

\[
W_{p,q} = \begin{cases} 
0, & \text{if } p \neq q \in \partial_1 \mathcal{P}, \\
\frac{1 + \omega(p) + \omega(q)}{2}, & \text{otherwise},
\end{cases}
\]

where \( \omega(\cdot) \) denotes the average pixel saliency of \( \mathcal{P} \) (see Fig. 2(d)). We normalize \( W_{p,q} \) in the range of \([1, 2]\) to avoid
over-penalizing saliency weights. As stitching results are usu-
ally cropped into rectangles in consumer-level photography,
we assign \( W_{p,q} = 0 \) if either \( p \) or \( q \) is located in the common
border \( \partial_1 \mathcal{P} \) of the canvas and \( \mathcal{P} \) (marked in green in Fig. 2(a)).

Finally, the perception-based energy function is defined as

\[
\tilde{E}(l) = \sum_{l \in \mathcal{P}} D_p(l_p) + \sum_{(p,q) \in \mathcal{N}} W_{p,q} \cdot \tilde{S}_{p,q}(l_p, l_q),
\]

where \( W_{p,q} \) rises the penalty of \( \tilde{S}_{p,q}(l_p, l_q) \) according to \( \omega(\cdot) \).
Fig. 2(f) shows that the endpoints of the seam have more freedom on \( \partial_1 \mathcal{P} \) than the seam shown in Fig. 2(c).

C. Proposed Seam-cutting Framework

Our seam-cutting framework is summarized in Algorithm 1.

III. EXPERIMENTS

In our experiments, first, we use SIFT \((27)\) to extract/match
features, use RANSAC \((28)\) to determine a global homography
and align input images. Then, for the overlapping region, we
use Ostu’s algorithm \((25)\) to estimate a threshold \( \tau (\epsilon = 0.06),\)
Algorithm 1 Perception-based seam-cutting framework.

**Input:** An overlapping region $P$ of aligned images $I_0$ and $I_1$.

**Output:** A stitching result.

1) Calculate $I_\ast(P)$ in Eq. (4);
2) Calculate $\tau$ in Eq. (5) via Ostu’s algorithm [25];
3) Calculate $I_l(P)$ in Eq. (7) and $S_{p,q}$ in Eq. (6);
4) Calculate $\omega(P)$ via salient object detection [26] and $W_{p,q}$ in Eq. (8);
5) Calculate $D_p(P)$ in Eq. (2);
6) Minimize $\tilde{E}(l)$ in Eq. (9) via graph-cut optimization [19], and blend $I_0$ and $I_1$ via gradient domain fusion [12].

and use salient object detection [26] to calculate pixel saliency weights. Finally, we use graph-cut optimization [19] to obtain a seam, and blend aligned images via gradient domain fusion [12] to create a mosaic.

Fig. 5 shows some experimental comparisons between two seam-cutting frameworks. Input images in the second group come from the dataset in [23]. Due to unconstrained shooting environment, there exist large parallax in these examples, such that a global homography can hardly align them. In such cases, the normal seam-cutting framework fails to produce artifact-free results, while our perception-based seam-cutting framework successfully creates plausible mosaics. More results and original input images are available in the supplementary material.

In order to investigate whether our proposed method is more consistent with human perception, we conduct a user study for comparing two seam-cutting frameworks. We invite 15 participants to rank 15 unannotated groups of stitching results (make a choice from 3 options: 1. A is better than B, 2. B is better than A, 3. A and B are even). Fig. 6 shows the user study result, which demonstrates that our stitching results win most users’ favor.

**IV. CONCLUSION**

In this paper, we propose a novel perception-based energy function in the seam-cutting framework, to handle image stitching challenges in consumer-level photography. Experiments show that our method outperforms the seam-cutting method of the normal energy function, and a user study demonstrates that our results are more consistent with human perception. In the future, we plan to generalize our method in the seam-driven framework to deal with image alignment.
REFERENCES

[1] R. Szeliski, “Image alignment and stitching: A tutorial,” Technical Report MSR-TR-2004-92, Microsoft Research, 2004.
[2] R. Szeliski and H.-Y. Shum, “Creating full view panoramic image mosaics and environment maps,” in Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, ser. SIGGRAPH ’97. ACM Press/Addison-Wesley Publishing Co., 1997, pp. 251–258.
[3] M. Brown and D. G. Lowe, “Automatic panoramic image stitching using invariant features,” Int. J. Comput. Vision, vol. 74, no. 1, pp. 59–73, 2007.
[4] J. Davis, “Mosaics of scenes with moving objects,” in Proceedings of the 14th Eur. Conf. Comput. Vision Pattern Recognit., 2014, pp. 2339–2346.
[5] A. Levin, A. Zomet, S. Peleg, and Y. Weiss, “Seamless image stitching of scenes with large motions and exposure differences,” in Proc. IEEE Conf. Comput. Vision Pattern Recognit., vol. 2, 2006, pp. 2498–2505.
[6] P. J. Burt and E. H. Adelson, “A multiresolution spline with application to image mosaics,” in Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques, ser. SIGGRAPH ’84. ACM Press/Addison-Wesley Publishing Co., 1984, pp. 306–313.
[7] J. Zaragoza, T.-J. Chin, M. S. Brown, and D. Suter, “As-projective-as-possible image stitching with moving DLT,” in Proc. IEEE Conf. Comput. Vision Pattern Recognit., 2013, pp. 2339–2346.
[8] J. Gao, S. J. Kim, and M. S. Brown, “Seam-driven image stitching,” in Eurographics and IEEE Conf. on Computer Graphics and Interactive Techniques, ser. SIGGRAPH ’01. ACM, 2001, pp. 341–346.
[9] A. A. Efros and W. T. Freeman, “Image quilting for texture synthesis and transfer,” in Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, ser. SIGGRAPH ’01. ACM, 2001, pp. 341–346.
[10] A. Mills and G. Dudek, “Image stitching with dynamic elements,” Image and Vision Computing, vol. 27, no. 10, pp. 1593–1602, 2009.
[11] Y. Boykov and V. Kolmogorov, “An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 2, pp. 147–159, Feb. 2004.
[12] J. Gao, Y. Li, T.-J. Chin, and M. S. Brown, “Seam-driven image stitching,” Eurographics, pp. 45–48, 2013.
[13] F. Zhang and F. Liu, “Parallax-tolerant image stitching,” in Proc. IEEE Conf. Comput. Vision Pattern Recognit., 2014, pp. 3262–3269.
[14] K. Lin, N. Jiang, L.-F. Cheong, M. Do, and J. Lu, “Seagull: Seam-guided local alignment for parallax-tolerant image stitching,” in Proc. 14th Eur. Conf. Comput. Vision, 2016, pp. 370–385.
[15] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen, “Interactive digital photomontage,” ACM Transactions on Graphics, vol. 23, no. 3, pp. 294–302, 2004.
[16] A. Eden, M. Uyttendaele, and R. Szeliski, “Seamless image stitching of scenes with large motions and exposure differences,” in Proc. IEEE Conf. Comput. Vision Pattern Recognit., vol. 2, 2006, pp. 2498–2505.
[17] J. Jia and C.-K. Tang, “Image stitching using structure deformation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 4, pp. 617–631, Apr. 2008.
[18] G. Zhang, Y. He, W. Chen, J. Jia, and H. Bao, “Multi-viewpoint panorama construction with wide-baseline images,” IEEE Transactions on Image Processing, vol. 25, no. 7, pp. 3099–3111, 2016.
[19] Y. Boykov, O. Veksler, and R. Zabih, “Fast approximate energy minimization via graph cuts,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 11, pp. 1222–1239, Nov. 2001.
[20] V. Kolmogorov and R. Zabin, “What energy functions can be minimized via graph cuts?” IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 2, pp. 147–159, Feb. 2004.
[21] J. Zaragoza, T.-J. Chin, M. S. Brown, and D. Suter, “As-projective-as-possible image stitching with moving DLT,” in Proc. IEEE Conf. Comput. Vision Pattern Recognit., 2013, pp. 2339–2346.
[22] N. Otsu, “A threshold selection method from gray-level histograms,” IEEE Trans. Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62–66, 1979.
[23] J. Davis, “Mosaics of scenes with moving objects,” in Proc. IEEE Conf. Comput. Vision Pattern Recognit., 1998, pp. 354–360.
[24] M. Brown and D. G. Lowe, “Automatic panoramic image stitching using invariant features,” Int. J. Comput. Vision, vol. 74, no. 1, pp. 59–73, 2007.
[25] M. Brown and D. G. Lowe, “Automatic panoramic image stitching using invariant features,” Int. J. Comput. Vision, vol. 74, no. 1, pp. 59–73, 2007.
[26] J. Zhang, S. Sclaroff, Z. Lin, X. Shen, B. Price, and R. Mech, “Minimum barrier salient object detection at 80 fps,” in Proc. IEEE Int. Conf. on Computer Vision, 2015, pp. 1404–1412.
[27] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” Int. J. Comput. Vision, vol. 60, no. 2, pp. 91–110, 2004.
[28] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” Commun. ACM, vol. 24, no. 6, pp. 381–395, 1981.
[29] R. Szeliski, “Interactive digital photomontage,” ACM Transactions on Graphics, vol. 23, no. 3, pp. 294–302, 2004.
[30] A. Eden, M. Uyttendaele, and R. Szeliski, “Seamless image stitching of scenes with large motions and exposure differences,” in Proc. IEEE Conf. Comput. Vision Pattern Recognit., vol. 2, 2006, pp. 2498–2505.
[31] J. Jia and C.-K. Tang, “Image stitching using structure deformation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 4, pp. 617–631, Apr. 2008.
[32] G. Zhang, Y. He, W. Chen, J. Jia, and H. Bao, “Multi-viewpoint panorama construction with wide-baseline images,” IEEE Transactions on Image Processing, vol. 25, no. 7, pp. 3099–3111, 2016.