POST-STITCHING DEPTH ADJUSTMENT FOR STEREOSCOPIC PANORAMA

Haoyu Wang⋆*, Daniel J. Sandin*†*, Dan Schonfeld*⋆

⋆ Department of Electrical and Computer Engineering, University of Illinois at Chicago
851 S. Morgan Street, Chicago, IL 60607, U.S.A.
† Electronic Visualization Laboratory, University of Illinois at Chicago
851 S. Morgan Street, Chicago, IL 60607, U.S.A.

ABSTRACT

In general, the viewing experience of stereoscopic panoramas heavily relies on the accuracy of perceived depth information. However, most of existed stereoscopic panorama stitching methods always focus on the improvement, refinement, and correction during the image alignment or panorama stitching. The impairment to the depth information introduced by later steps is rarely under consideration. Thus, we propose one general post-stitching depth correction technique based on sparsely detected control points. The proposed depth correction can be divided into two steps. The global one aims at minimizing the average depth error by adjusting the relative poses of two panoramas. Then, the local step tries to relieve those regions-based depth issues with the morphing technique. For those regions that still suffer from visible depth error after the global correction, we select control points in those depth anomaly patches and utilize the Thin-Plate-Spline technique to warp those features into their target position. To demonstrate the robustness of the proposed depth adjustment technique, various experiments are conducted under different camera arrangements and scenarios. The result indicates that the stereoscopic panoramas after proposed depth adjustment could deliver more reasonable depth information and bring a better viewing experience.

Index Terms— Stereoscopic Panorama, Commonly-identified Feature, Visual Saliency Index, Target Disparity Map, Thin-Plate-Spline

1. INTRODUCTION

As the extensive investigations in high-quality stereoscopic panorama generation and widespread usage of VR display equipment, the comfortable immersive visual experience of real-world scenes are always expected by the audiences. The criterion for stitching quality of stereoscopic panoramas not only includes the misalignment, stitching errors and object distortion but also relies on the accuracy of the depth information. Although many complicated techniques and hardware-orientated solutions are proposed to handle the depth control problem [1, 2, 3, 4], most of the correction, refinement, and adjustment are operated before or in the panorama generation process. Those later introduced depth disturbing, such as inconsistent blending seams and panorama straighten, are always ignored. Thus, the goal of this study is to provide an efficient general post-stitching depth correction strategy that could minimize depth error and stereo inconsistency with sparsely sampled depth information.

This paper presents a two-step depth correction strategy to correct the perceived depth into a comfortable range. Based on the well-matched commonly-identified feature set from input images, we firstly operate the global translation to adjust the relative pose between left and right view panorama. The fitted translation vector, which causes the minimal stereo inconsistency, can ensure the majority of pixels in the stitched panoramas correctly deliver depth information. Then, we utilize Thin-Plate-Spline warping method to fix all the noticeable depth error in small regions after the global correction, according to the target disparity map from the input rectified image.

2. RELATED WORKS

Traditional monocular panorama stitching methods cannot handle stereo consistency well due to its lack of depth information utilization [5, 6, 7]. Therefore, recently proposed stereoscopic panorama generation methods provide various solutions to alleviate the stereo visual discomfort caused by the inaccurate depth information [3, 4, 8, 9]. Based on the rotation of cameras in a circular trajectory or static radial camera array, the omnistereo projection method [1] and it extended methods [4, 10, 11, 12] usually implement depth adjustment operation by careful selection of corresponding left and right view strips. While Richardt et al. [2] proposed to make compensation to the vertical disparity before stitching by projecting undistorted input images onto a cylindrical imaging surface. Zhang and Liu also extended a spatially varying warping method [13] to warp the input images under the guidance of well-stitched dense disparity map [3]. However, all above depth control strategy is highly correlated to its own unique hardware setup and stitching algorithm, which indicates the difficulty in generalization and extension. Thus, we intend to propose one general depth correction strategy that can fit different camera arrangements, captured scenarios and panorama generation methods. The whole depth adjustment process can only depend on the input rectified image pairs and originally stitched stereoscopic panoramas.

3. PROPOSED DEPTH ADJUSTMENT

3.1. Global Depth Adjustment

The first correction step can be interpreted as global registration between the left and right panorama. We wish to adjust the output panoramas with a translation vector $< d_v, d_h >$ for better global depth perception. For simplicity, we explain its details in the stitching task for only two pairs of input images, $\{L_1, L_2, R_1, R_2\}$. The basic unit we used for global depth correction is called commonly-identified feature(CIF), which refers to the same corner, edge, or region observed and precisely described by all of the adjacent camera views. The technique for detection and construction of the CIF in [14] is directly utilized here. Thus, for the two pairs of images, we can produce one corresponding commonly-identified feature set $S = \{d_{i,1}, d_{i,2}, d_{i,3}, d_{i,4}; i = 1 : N\}$. In the ideal case, those features' final projection position in the output panorama is expected to provide identical depth information as what we perceive from the input images. Then, two corresponding stereo consistency errors can be defined as:

$$E_v(i) = |d'_{i,1}.y - d'_{i,3}.y| + |d'_{i,2}.y - d'_{i,4}.y| \tag{1}$$

$$\begin{aligned} E_h(i) = &|(d'_{i,1}.x - d'_{i,3}.x) - (d_{i,1}.x - d_{i,3}.x)| \\ &+ |(d'_{i,2}.x - d'_{i,4}.x) - (d_{i,2}.x - d_{i,4}.x)|. \end{aligned} \tag{2}$$

In the above two equations, $d_{i,1}$, $d_{i,2}$, $d_{i,3}$, and $d_{i,4}$ are the ith four matched features in the commonly identified set S. The primed symbols $d'_{i,1}$, $d'_{i,2}$, $d'_{i,3}$, and $d'_{i,4}$ are their corresponding features in the generated panorama. Additionally, $d_{i,1}.x$ and $d_{i,1}.y$ represent the feature's center point position.

Thus, the stereo consistency errors after the translation operation with d_v and d_h turn to be:

$$E_v^g(i, d_v) = |d'_{i,1}.y + d_v - d'_{i,3}.y| + |d'_{i,2}.y + d_v - d'_{i,4}.y| \tag{3}$$

$$\begin{aligned} E_h^g(i, d_h) = &|(d'_{i,1}.x + d_h - d'_{i,3}.x) - (d_{i,1}.x - d_{i,3}.x)| \\ &+ |(d'_{i,2}.x + d_h - d'_{i,4}.x) - (d_{i,2}.x - d_{i,4}.x)|. \end{aligned} \tag{4}$$

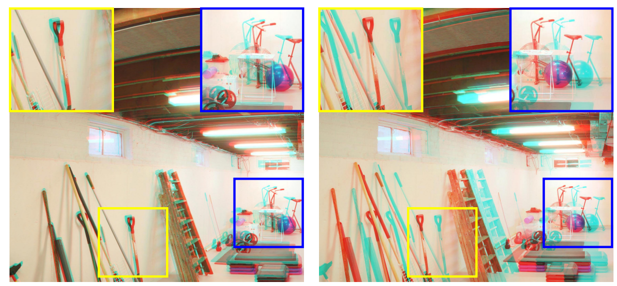
The global depth correction can be formulated as the optimization problem to fit two motion scalars $\hat{d_v}$ and $\hat{d_h}$ that cause the minimal stereo consistency errors for all commonly identified features in set S:

$$\hat{d_v} = arg\min_{d \in \mathcal{R}} \sum_{i=1}^{N} v_i E_v^g(i, d_v) \tag{5}$$

$$\hat{d_h} = arg\min_{d \in \mathcal{R}} \sum_{i=1}^{N} v_i E_h^g(i, d_h) \tag{6}$$

Weight v_i indicate visual saliency index of the corresponding commonly-identified feature [15, 16], which characterize the visual importance of features. The corporation of saliency weights can force the fitted translation care more about those features which attract more attention from viewers.

Two cropped stereoscopic panoramas in red-cyan anaglyph are stated in Figure 1. In the right panorama, we can see the vertical disparity issue after the global translation is nearly eliminated. The horizontal disparity of bicycle and shovel are also adapt to one reasonable range that delivers correct depth information.

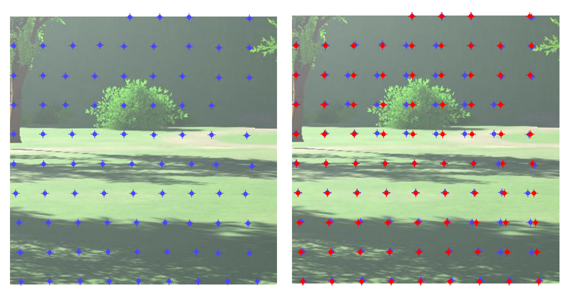


(a) Before Global Correction (b) After Global Correction

Fig. 1. Comparison before and after global depth correction

3.2. Local Depth Adjustment

Whereas the global correction can largely relieve the viewing discomfort, we can always find some tiny artifacts in its corrected result. To remove these undesirable issues, one Thin-Plate-Spline(TPS) based morphing method is proposed to fix the region of the interest under the guidance of target disparity map. Without losing the generality, we consider the right view panorama with better monocular stitching quality as the reference and perform TPS warping at left view panorama in the following discussion.



(a) Original position (b) Expected position

Fig. 2. expected position of sampled control points

3.2.1. Control Point Generation

Operation of TPS warping requires two equally sized corresponding point-sets in the areas with depth anomaly. The region of the interest is usually manually labeled for depth correction and denoted as G_L and G_R respectively. Afterward, one group of control points $\{P^L_{i,j}\}$ can be detected and extracted from the left view ROI, where i is the index for the sampled pixel and j is the index for ROI. Given the projection matrix H, which describes the geometrical transformation from the camera view to the panorama view, we utilize its inverse function to obtain the position of the corresponding point at the input rectified image coordinate. According to the target disparity map from input images, the expected disparity of those sampled control points can be computed easily. Thus, the expected position of control point after the local warping can be defined as:

$$\hat{P}^L_{i,j}.x = P^R_{i,j}.x + Disp(H^{-1}(P^L_{i,j}.x, P^L_{i,j}.y)) \times R \qquad (7)$$

$$\hat{P}^L_{i,j}.y = P^R_{i,j}.y. \qquad (8)$$

In the above equations, $P^L_{i,j}.x$, $P^L_{i,j}.y$, $P^R_{i,j}.x$ and $P^R_{i,j}.y$ represent the position of selected control point in the left and right view panorama. The hatted symbols $\hat{P}^L_{i,j}.x$ and $\hat{P}^L_{i,j}.y$ refer to the expected position of the sampled control points in the left view panorama. Besides, H^{-1} is the inverse of projection matrix, $Disp$ is the disparity map from rectified image pair and R is the ratio of pixel per degree between panorama view and camera view.

In the example depicted in Figure 2, the blue stars in the fist sub-figure mark the position of the sampled control points and the red stars in the second sub-figure indicate their expected position with correct depth.

3.2.2. Depth Awared TPS Warping

The standard TPS [17] can fit one mapping function, Φ, between the two equally sized corresponding point-sets, $A = \{x_a, y_a\}$ and $B = \{x_b, y_b\}$, with minimal bending energy:

$$E_{tps}(\Phi) = \sum_{i=1}^{M} ||v_i - \Phi(x_{a,i}, y_{a,i})||^2 + \iint_R [(\frac{\partial^2 \Phi}{\partial x^2}) + 2(\frac{\partial^2 \Phi}{\partial x \partial y})^2 + (\frac{\partial^2 \Phi}{\partial y^2})]\, dx\, dy \qquad (9)$$

In our case, we set v_i equal to the target coordinates (x_b, y_b) in turn to obtain two continuous transformation for x and y coordinate respectively. Point set A is defined as the original control point $P^L_{i,j}$ and B refers to the corresponding point $\hat{P}^L_{i,j}$ at expected position. According to the proof in reference [18], the unique minimizer, Φ, is parameterized as follows:

$$\Phi(x, y) = \gamma_1 + \gamma_2 * x + \gamma_3 * y + \sum_{i=1}^{M} w_i U(|(x_{a,i}, y_{a,i}) - (x, y)|) \qquad (10)$$

$$U(r) = \begin{cases} r^2 \ln r & r > 0 \\ 0 & r = 0 \end{cases} \qquad (11)$$

Therefore, we intend to find the coefficients $[w|\gamma_1, \gamma_2, \gamma_3]$ in mapping function Φ. In order for Φ to have square integrable at second derivatives, we require that

$$\sum_{i=1}^{M} w_i = \sum_{i=1}^{M} w_i x_{a,i} = \sum_{i=1}^{M} w_i y_{a,i} = 0. \qquad (12)$$

Together with the exact interpolation conditions, $\Phi(x_{a,i}, y_{a,i}) = v_i$, this produces a linear system as follows:

$$\begin{bmatrix} \boldsymbol{K} & \boldsymbol{P} \\ \boldsymbol{P}^T & \boldsymbol{O} \end{bmatrix} \begin{bmatrix} \boldsymbol{w} \\ \boldsymbol{\gamma} \end{bmatrix} = \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{0} \end{bmatrix}, \qquad (13)$$

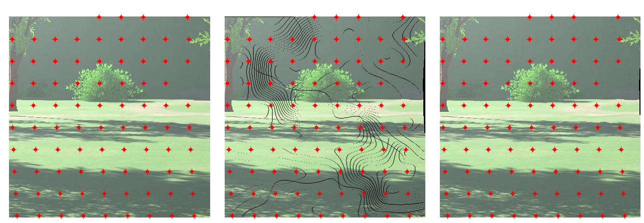
where $\boldsymbol{K}$, $\boldsymbol{P}$, and $\boldsymbol{O}$ are submatrices

$$K_{i,j} = U(|(x_{a,i}, y_{a,i}) - (x_{a,j}, y_{a,j})|)$$
$$\boldsymbol{P}_{M\times 3} = \begin{bmatrix} 1 & x_{a,1} & y_{a,1} \\ 1 & x_{a,2} & y_{a,2} \\ \vdots & \vdots & \vdots \\ 1 & x_{a,M} & y_{a,M} \end{bmatrix} \qquad (14)$$
$$\boldsymbol{O}_{3\times 3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

and $\boldsymbol{w}, \boldsymbol{\gamma}$, and $\boldsymbol{v}$ are column vectors, which stand for TPS coefficients and target data value respectively Then, we can obtain the TPS interpolation coefficients as:

$$\begin{bmatrix} \boldsymbol{w} \\ \boldsymbol{\gamma} \end{bmatrix} = \begin{bmatrix} \boldsymbol{K} & \boldsymbol{P} \\ \boldsymbol{P}^T & \boldsymbol{O} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{0} \end{bmatrix}. \qquad (15)$$

Once TPS coefficients $[w|\gamma_1, \gamma_2, \gamma_3]$ is computed, we use the TPS equation (10) to find the expected position for those unsampled points in the ROI. After all pixels in the ROI have been projected into their new position via the TPS coefficients, we will obtain the adjusted ROI with correct depth information. In Figure 3, sub-figure (a) is the unwrapped ROI with control points at their expected position. Then, sub-figure (b) shows the warped ROI after projection of all pixel under TPS coefficients. Finally, the sub-figure (c) is the warped ROI after all black holes have been filled via nearest neighbor interpolation.

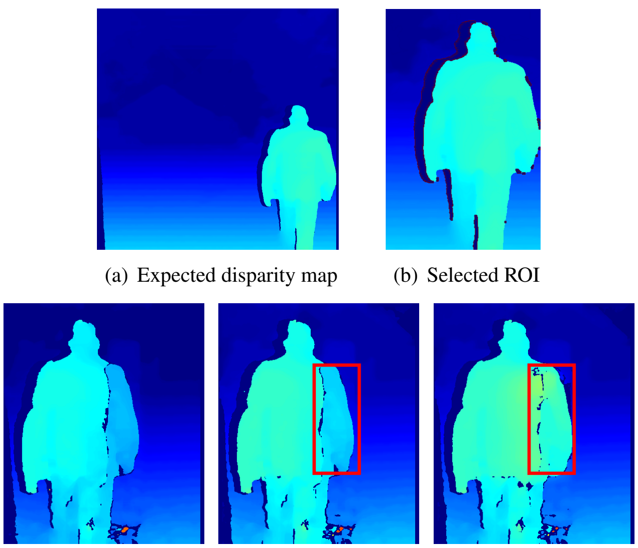


(a) (b) (c)

Fig. 3. Thin-Plate-Spline warping

	Horizontal Dist		Vertical Dist	
	Before	After	Before	After
Atrium	1.62px	1.41px	1.15px	1.12px
Basement	2.69px	0.67px	1.89px	0.84px
Campus	2.13px	2.02px	1.53px	1.53px
Rampart	28.87px	1.81px	1.79px	1.19px
Barcelona	1.96px	1.54px	2.30px	2.23px
Classroom	1.30px	1.27px	3.89px	1.58px
Village-a	1.62px	1.41px	1.15px	1.12px
Village-b	1.40px	1.29px	1.21px	1.11px
Village-c	1.35px	1.30px	1.05px	1.02px
Living-room-a	1.09px	0.91px	1.96px	1.94px
Living-room-b	0.68px	0.59px	1.73px	1.65px

Table 1. Depth error before and after global correction



(a) Expected disparity map (b) Selected ROI

(c) Before correction (d) Globally corrected (e) Locally corrected

Fig. 4. Disparity Map before and after correction

4. EXPERIMENTS

4.1. Experiment Setup

To validate the effectiveness and robustness of our proposed depth correction technique, we employ the feature-based stereoscopic panorama stitching framework [19] as the baseline. In the panorama generation process, $vlfeat$ [20] lib is used for SIFT detection and $Enblend$ [21] is the panorama blender. All stitched panoramas are scaled to 12000 by 6000 pixels for $360° \times 180°$. Each local region is manually labeled as 400 by 400-pixel rectangle and there are 400 control points uniformly sampled for each local region warping.

4.2. Visual Comparison

One instance in Figure 4 visually demonstrates improvement to the perceived disparity map. Sub-figure (a) is the full-size target disparity map estimated from input rectified image pairs, and sub-figure (b) is the cropped version for the selected ROI. Sub-figure (c), (d) and (e) shows the measured disparity map of ROI before global correction, after global correction and after local correction, respectively. It's evident that the global correction shifts the overall disparity value of ROI into the range more similar to the target map. Besides, the perceived disparity value of the human's right arm, which is marked with the red rectangle, is also corrected.

	Horizontal Dist		Vertical Dist	
	Before	After	Before	After
Atrium	2.78px	1.93px	1.13px	1.12px
Basement	3.40px	1.10px	1.10px	0.89px
Village-a	1.95px	0.65px	1.08px	1.05px
Village-b	1.70px	0.63px	1.23px	1.04px
Village-c	1.85px	0.57px	1.10px	1.08px
Living-room-a	2.70px	1.58px	2.03px	1.94px
Living-room-b	2.10px	1.47px	1.87px	1.86px

Table 2. Depth error in ROIs before and after local correction

4.3. Numerical Comparison

Both of camera-captured and synthetic data are tested to quantify the improvement of our proposed depth correction to the original stitching result. Considering disparity map from input rectified image pairs as ground truth and uniformly sampled features as testing control points, the difference between the perceived depth of testing control points from stereoscopic panoramas and expected depth from ground truth is then used as the metric to evaluate the performance of depth adjustment. The pixel-level depth error for global and local correction are stated in Table 1 and Table 2 respectively. The testing data-set includes four frames of real-captured outdoor scenarios, two frames of synthetic indoor scenarios and five animations. Missing of several static camera-based data-sets in Table 2 indicate no local depth anomaly region found. The depth error recorded in tables is the average of 30 frame panorama in each animation data-set, while sometimes the pixel-wise metric fails to characterize the improvement for human visual perception to depth.

5. CONCLUSION

In this paper, we presented a general depth correction strategy in the stereoscopic panorama generation system. Given the well-matched commonly-identified features from input image pairs, we consider one of the stitched stereoscopic panoramas as the reference and globally shift another one for minimal average depth error globally. Those areas with local noticeable depth issues are then labeled and fixed with Thin-Plate-Spline warping under the guidance of sparse target disparity map. Extensive simulations under different scenarios are tested to prove the effectiveness and robustness of our method.

6. ACKNOWLEDGMENT

This publication is based on work supported in part by the National Science Foundation award CNS-1456638 for SENSEI.

7. REFERENCES

[1] Shmuel Peleg, Moshe Ben-Ezra, and Yael Pritch, "Omnistereo: Panoramic stereo imaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 279–290, 2001.

[2] Christian Richardt, Yael Pritch, Henning Zimmer, and Alexander Sorkine-Hornung, "Megastereo: Constructing high-resolution stereo panoramas," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1256–1263.

[3] Fan Zhang and Feng Liu, "Casual stereoscopic panorama stitching," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2002–2010.

[4] Robert Anderson, David Gallup, Jonathan T Barron, Janne Kontkanen, Noah Snavely, Carlos Hernández, Sameer Agarwal, and Steven M Seitz, "Jump: virtual reality video," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 6, pp. 198, 2016.

[5] Richard Szeliski et al., "Image alignment and stitching: A tutorial," *Foundations and Trends® in Computer Graphics and Vision*, vol. 2, no. 1, pp. 1–104, 2007.

[6] Matthew Brown and David G Lowe, "Automatic panoramic image stitching using invariant features," *International journal of computer vision*, vol. 74, no. 1, pp. 59–73, 2007.

[7] Junhong Gao, Yu Li, Tat-Jun Chin, and Michael S Brown, "Seam-driven image stitching.," in *Eurographics (Short Papers)*, 2013, pp. 45–48.

[8] Ho-Chao Huang and Yi-Ping Hung, "Panoramic stereo imaging system with automatic disparity warping and seaming," *Graphical Models and Image Processing*, vol. 60, no. 3, pp. 196–208, 1998.

[9] Federico Perazzi, Alexander Sorkine-Hornung, Henning Zimmer, Peter Kaufmann, Oliver Wang, S Watson, and M Gross, "Panoramic video from unstructured camera arrays," in *Computer Graphics Forum*. Wiley Online Library, 2015, vol. 34, pp. 57–68.

[10] Vincent Couture, Michael S Langer, and Sebastien Roy, "Panoramic stereo video textures," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1251–1258.

[11] Vincent C Couture, Michael S Langer, and Sebastien Roy, "Omnistereo video textures without ghosting," in *3D Vision-3DV 2013, 2013 International Conference on*. IEEE, 2013, pp. 64–70.

[12] Chiao Wang and Alexander A Sawchuk, "Region-based stereo panorama disparity adjusting," in *Multimedia Signal Processing, 2006 IEEE 8th Workshop on*. IEEE, 2006, pp. 186–191.

[13] Feng Liu, Michael Gleicher, Hailin Jin, and Aseem Agarwala, "Content-preserving warps for 3d video stabilization," *ACM Transactions on Graphics (TOG)*, vol. 28, no. 3, pp. 44, 2009.

[14] Haoyu Wang, Daniel J Sandin, and Dan Schonfeld, "Saliency-based feature selection strategy in stereoscopic panoramic video generation," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 1837–1841.

[15] Jonathan Harel, Christof Koch, and Pietro Perona, "Graph-based visual saliency," in *Advances in neural information processing systems*, 2007, pp. 545–552.

[16] Xiaodi Hou, Jonathan Harel, and Christof Koch, "Image signature: Highlighting sparse salient regions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 1, pp. 194–201, 2012.

[17] Fred L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 11, no. 6, pp. 567–585, 1989.

[18] Grace Wahba et al., "Spline models for observational data. society for industrial and applied mathematics," 1990.

[19] Haoyu Wang, Daniel J Sandin, and Dan Schonfeld, "A common feature-based disparity control strategy in stereoscopic panorama generation," in *Visual Communications and Image Processing (VCIP), 2017 IEEE*. IEEE, 2017, pp. 1–4.

[20] Andrea Vedaldi and Brian Fulkerson, "Vlfeat: An open and portable library of computer vision algorithms," in *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 2010, pp. 1469–1472.

[21] Andrew Mihal, P dAngelo, et al., "Enblend," 2012.