depth or *disparity map* as intermediate product. Section 2.1.1 briefly discusses the fundamentals of the stereoscopic computation of disparity maps from stereoscopic videos.

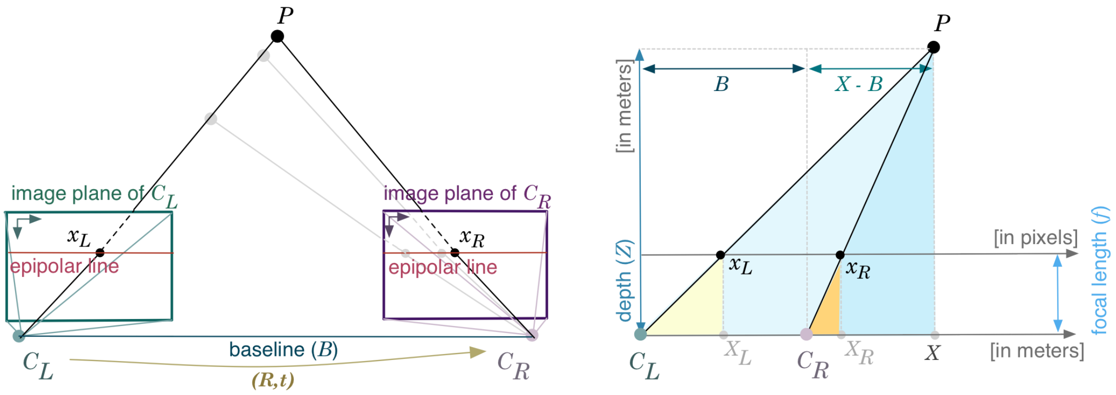
In many applications, the disparity map is computed from two views of the same scene using stereo vision approaches. However, if only one view is available, such as for existing monocular videos, 2D-to-3D conversion approaches can be considered as an alternative solution. The conversion can be performed manually by assigning disparities[1] to each pixel of a video, semi-automatically by propagating sparse user-given disparities over the entire video, or fully-automatically by investigating monocular depth cues [119, 141]. Like in the stereo case, disparity maps that were computed by these approaches can be adjusted to different types of displays and utilized for novel view generation. Section 2.1.2 provides a general overview of different types of 2D-to-3D conversion approaches and their principles. In Section 2.2, special emphasis is put on the state-of-the-art of semi-automatic 2D-to-3D conversion, which is the focus of this thesis.

Additional approaches for generating content for 3D viewing may be based on the availability of a 3D model, from which disparities and multiple views can be rendered, e.g., with 3D computer graphic software such as Blender [10]. Furthermore, depth can be captured directly with special depth sensors and scanners such as Microsoft Kinect [103].

2.1.1 3D from Stereoscopic Data

Given multiple, e.g., two, images that were taken from slightly shifted viewpoints of the same scene, a 3D model of the scene can be estimated by determining pixel correspondences between these images (*stereo correspondence problem* or *stereo matching problem* [146]). The shift in position of these corresponding pixels, the *disparity*, directly relates to the depth of a scene. This relationship can be derived from the *standard rectified stereo geometry* [146]. In particular, Figure 2.1 *a)* illustrates the two images within the standard rectified camera setup captured by two cameras. The cameras C_L and C_R are connected by a horizontal line, which is called the *baseline*. C_L and C_R are *calibrated*, i.e., the transformation (R, t) of the camera coordinate system of one camera to the other camera is known, and *rectified*, i.e., the image planes of C_L and C_R lie in a common plane that is parallel to the baseline. (For more details concerning camera calibration and rectification interested readers are referred to [146].) When capturing a 3D scene using the camera setup in Figure 2.1 *a)*, the 3D point P is projected into the points x_L and x_R on the image planes of C_L and C_R. During this process, C_L, C_R, P, x_L and x_R span a plane, the *epipolar plane* [146]. Due to the rectified camera setup, matching points in one image plane (e.g., x_L in the left view and x_R in the right view) must lie on a particular horizontal line that intersects the epipolar plane with the image plane, i.e., the corresponding *epipolar line*, in the other view. This restriction concerning the location of corresponding points provides an advantageous *epipolar*

[1]The term disparity was introduced to describe position differences in stereoscopic conditions and refers to the field of stereo vision [101, 146]. However, semi-automatic 2D-to-3D conversion algorithms (e.g., [56, 117, 163]) use equivalent values for their depth information. As stereo disparity, it encodes the closeness of pixels to the camera (i.e., is large in the fore- and low in the background) and can be used to generate novel views by shifting pixel positions accordingly. The depth information used in 2D-to-3D conversion algorithms is typically either, exactly as disparity, given in terms of position shifts that can be used directly to generate novel views (e.g., [56, 163]) or as normalized values $\in [0, 1]$ that have to be scaled prior to that (e.g., [117]). It is not given in meters as scene depths. In the 2D-to-3D conversion literature the terms disparity and depth are used both. As in [56, 163], in this thesis we use the term disparity when referring to depth information in the context of 2D-to-3D conversion.



a) b)

Figure 2.1: Standard rectified epipolar geometry. *a)* The horizontally neighboring cameras C_L and C_R are *rectified*, i.e., the image planes lie in a common plane that is parallel to the *baseline*. Matching points in one view lie on a horizontal line, i.e., the *epipolar line*, in the other view. C_L and C_R are *calibrated*, i.e., the transformation (R,t) between their camera coordinate systems is known. *b)* In this setup the *disparity* $d_x = x_R - x_L$ and depth Z of a 3D point P with coordinates (X,Y,Z) and its projections in the image planes with x_L and x_R are related via similar triangles (i.e., (x_L, X_L, C_L), (P, X, C_L), (x_R, X_R, C_R) and (P, X, C_R)). [11, 146]

constraint, which reduces the search space for corresponding pixels in the left and the right view to their horizontal scan-lines.

Having identified two corresponding pixels, e.g., x_L and x_R, which are located in the left and the right view, the *disparity* d_x can be determined by their horizontal position shift, i.e., $d_x = x_L - x_R$. As shown in Figure 2.1 *b)*, the disparity d_x is inversely proportional to the depth Z of a scene. They are related via the similar triangles (x_L, X_L, C_L), (P, X, C_L), (x_R, X_R, C_R) and (P, X, C_R), which leads to the following equation:

$$Z = f\frac{B}{d_x}. \tag{2.1}$$

Here, f is the focal length (in pixels) and B is the baseline between C_L and C_R. Thus, the task of estimating depth from a stereo image pair is reduced to the task of estimating the disparity of each pixel (*disparity map)*. In the context of the standard rectified stereo geometry, the process of stereo matching can be solved by finding corresponding (*matching*) pixels in horizontal scan-lines of the left and the right view. A stereo matching algorithm's foundation to find these correspondences is the definition of a measure that expresses the quality (or *matching costs*) of a potential match between a pixel of the left and a pixel of the right view. This is typically done by measuring the similarity, e.g., the color difference, of these pixels [134, 146]. While high similarities indicate good matches, large matching costs point to a low matching quality. As a second step, these costs can be aggregated. The final pixel correspondences (and thus the resulting disparity map) are determined in terms of an optimization that is defined over the previously computed costs. This optimization can be performed *locally* (e.g., [125, 126]), by selecting the disparities with the lowest costs according to a local pixel neighborhood or *globally* (e.g., [12–14]), by minimizing a

Bibliography

[1] A. Agarwala, A. Hertzmann, D. H. Salesin, and S. M. Seitz. Keyframe-based tracking for rotoscoping and animation. In *SIGGRAPH'04*, pages 584–591, 2004.

[2] S. Aguirre and R. M. Rodriguez-Dagnino. Synthesizing stereo 3D views from focus cues in monoscopic 2D images. In *EI'03: Electronic Imaging, Stereoscopic Displays and Virtual Reality Systems X*, pages 377–388, 2003.

[3] J. Bai and X. Wu. Error-tolerant scribbles based interactive image segmentation. In *CVPR'14: Computer Vision and Pattern Recognition*, pages 392–399, 2014.

[4] X. Bai and G. Sapiro. Geodesic Matting: A framework for fast interactive image and video segmentation and matting. *International Journal of Computer Vision*, 82(2):113–132, 2009.

[5] X. Bai, J. Wang, D. Simons, and G. Sapiro. Video SnapCut: Robust video object cutout using localized classifiers. In *SIGGRAPH'09*, pages 70:1–70:11, 2009.

[6] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. In *ICCV'07: International Conference on Computer Vision*, pages 1–8, 2007.

[7] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92(1):1–31, 2011.

[8] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen. Interactively co-segmentating topically related images with intelligent scribble guidance. *International Journal of Computer Vision*, 93(3):273–292, 2011.

[9] R. Bellman. On a routing problem. Technical report, DTIC Document, 1956.

[10] Blender. http://www.blender.org/. Accessed: 2014-05-28.

[11] M. Bleyer. *Segmentation-based Stereo and Motion with Occllusions*. PhD thesis, Vienna University of Technology, Institute of Software Technology and Interactive Systems, 2006.

[12] M. Bleyer and M. Gelautz. Simple but effective tree structures for dynamic programming-based stereo matching. In *VISAPP'08: International Conference on Computer Vision Theory and Applications*, pages 415–422, 2008.

[13] M. Bleyer, M. Gelautz, C. Rother, and C. Rhemann. A stereo approach that handles the matting problem via image warping. In *CVPR:'09: Conference on Computer Vision and Pattern Recognition*, pages 501–508, 2009.

[14] M. Bleyer, C. Rother, P. Kohli, D. Scharstein, and S. Sinha. Object Stereo - Joint stereo matching and object segmentation. In *CVPR'11: Conference on Computer Vision and Pattern Recognition*, pages 3081–3088, 2011.

[15] A. Bokov, D. Vatolin, A. Zachesov, A. Belous, and M. Erofeev. Automatic detection of artifacts in converted S3D video. In *EI'14: Electronic Imaging, Stereoscopic Displays and Applications XXV*, pages 1–14, 2014.

[16] J. S. Boreczky and L. A. Rowe. Comparison of video shot boundary detection techniques. *Journal of Electronic Imaging*, 5(2):122–128, 1996.

[17] G. Borgefors. Distance transformations in digital images. *Computer Vision, Graphics, and Image Processing*, 34(3):344–371, 1986.

[18] Y. Boykov and G. Funka-Lea. Graph cuts and efficient N-D image segmentation. *International Journal of Computer Vision*, 70(2):109–131, 2006.

[19] Y. Boykov and M. P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In *ICCV'01: International Conference on Computer Vision*, pages 105–112, 2001.

[20] Broadband Network & Digital Media Lab. Test sequences. http://media.au.tsinghua.edu.cn/2Dto3D/testsequence.html. Accessed: 2013-10-22, 2013.

[21] N. Brosch, A. Hosni, G. Ramachandran, L. He, and M. Gelautz. Content generation for 3D video/TV. *e & i Elektrotechnik und Informationstechnik*, 128(10):359–365, 2011.

[22] N. Brosch, A. Hosni, C. Rhemann, and M. Gelautz. Spatio-temporally coherent interactive video object segmentation via efficient filtering. In *DAGM/ÖAGM'12: Joint Symposium of the German Association for Pattern Recognition and the Austrian Association for Pattern Recognition*, pages 418–427, 2012.

[23] N. Brosch, M. Nezveda, M. Gelautz, and F. Seitner. Efficient quality enhancement of disparity maps based on alpha matting. In *EI'14: Electronic Imaging, Stereoscopic Displays and Applications XXV*, pages 1–10, 2014.

[24] N. Brosch, C. Rhemann, and M. Gelautz. Segmentation-based depth propagation in videos. In *ÖAGM'11: Austrian Association for Pattern Recognition Workshop*, pages 1–8, 2011.

[25] N. Brosch, T. Schausberger, and M. Gelautz. Towards perceptually coherent depth maps in 2D-to-3D conversion. In *EI'16: Electronic Imaging, Stereoscopic Displays and Applications XXVII*, pages 1–11, 2016.

[26] T. Brox and J. Malik. Object segmentation by long term analysis of point trajectories. In *ECCV'10: European Conference on Computer Vision: Part V*, pages 282–295, 2010.

[27] D. Butler, J. Wulff, G. B. Stanley, and M. J. Black. A naturalistic open source movie for optical flow evaluation. In *ECCV'12: European Conference on Computer Vision*, pages 611–625, 2012.

[28] X. Cao, Z. Li, and Q. Dai. Semi-automatic 2D-to-3D conversion using disparity propagation. *Transactions on Broadcasting*, 57(2):491–499, 2011.

[29] J. Chen, J. Zhao, X. Wang, C. Huang, E. Dong, B. Chen, and Z. Yuan. A simple semi-automatic technique for 2D to 3D video conversion. In *AICI'09: Artificial Intelligence and Computational Intelligence*, pages 336–343, 2011.

[30] X. Chen, D. Zou, J. Li, X. Cao, Q. Zhao, and H. Zhang. Sparse dictionary learning for edit propagation of high-resolution images. In *CVPR'14: Conference on Computer Vision and Pattern Recognition*, pages 2854–2861, 2014.

[31] J. W. Choi and T. K. Whangbo. A key frame-based depth propagation for semi-automatic 2D-to-3D video conversion using a pair of bilateral filters. *International Journal of Digital Content Technology*, 7(17):94–103, 2013.

[32] C. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

[33] C. Cotsaces, N. Nikolaidis, and I. Pitas. Video shot detection and condensed representation. A review. *Signal Processing Magazine*, 23(2):28–37, 2006.

[34] T. Cour, F. Benezit, and J. Shi. Spectral segmentation with multiscale graph decomposition. In *CVPR'05: Computer Vision and Pattern Recognition*, pages 1124–1131, 2005.

[35] F. C. Crow. Summed-area tables for texture mapping. In *SIGGRAPH'84*, pages 207–212, 1984.

[36] D. DeMenthon, X. Munoz, D. Raba, J. Marti, and X. Cufi. Spatio-temporal segmentation of video by hierarchical mean shift analysis. In *SMVP'02: Statistical Methods in Video Processing Workshop*, pages 142–151, 2002.

[37] X. Ding, Y. Xu, L. Deng, and X. Yang. Colorization using quaternion algebra with automatic scribble generation. In *MMM'12: Advances in Multimedia Modeling*, pages 103–114, 2012.

[38] D. Donatsch, N. Farber, and M. Zwicker. 3D conversion using vanishing points and image warping. In *3DTV-CON'13: 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video*, pages 1–4, 2013.

[39] Z. Farbman, F. Fattal, D. Lischinski, and R. Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. In *SIGGRAPH'08*, pages 67:1–67:10, 2008.

[40] Z. Farbman, R. Fattal, and D. Lischinski. Diffusion maps for edge-aware image editing. In *SIGGRAPH Asia'10*, pages 145:1–145:10, 2010.

[41] M. Fawaz, R. Phan, R. Rzeszutek, and D. Androutsos. Adaptive 2D to 3D image conversion using a hybrid graph cuts and random walks approach. In *ICASSP'12: International Conference on Acoustics, Speech and Signal Processing*, pages 1441–1444, 2012.

[42] C. Fehn, K. Schüür, I. Feldmann, P. Kauff, and A. Smolic. Distribution of ATTEST test sequences for EE4 in MPEG 3DAV. MPEG Meeting, 2002.

[43] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59:167–181, 2004.

[44] J. Feng, H. Ma, J. Hu, L. Cao, and H. Zhang. Superpixel based depth propagation for semi-automatic 2D-to-3D video conversion. In *ICNDC'12: International Conference on Networking and Distributed Computing*, pages 157–160, 2012.

[45] O. P. Gangwal and R. P. Berretty. Depth map post-processing for 3D-TV. In *ICCE'09: International Conference on Consumer Electronics*, pages 1–2, 2009.

[46] E. S. L. Gastal and M. Oliveira. Domain transform for edge-aware image and video processing. In *SIGGRAPH'11*, pages 69:1–69:12, 2011.

[47] S. Ghuffar, N. Brosch, N. Pfeifer, and M. Gelautz. Motion segmentation in videos from time of flight cameras. In *IWSSIP'12: Conference on International Systems, Signals and Image Processing*, pages 328–332, 2012.

[48] S. Ghuffar, N. Brosch, N. Pfeifer, and M. Gelautz. Motion estimation and segmentation in depth and intensity videos. *Integrated Computer-Aided Engineering*, 21(3):203–218, 2014.

[49] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. *Transactions on Pattern Analysis and Machine Intelligence*, 34(10):1915–1926, 2012.

[50] G. H. Golub and C. F. Van Loan. *Matrix computations*, volume 3. JHU Press, 2012.

[51] L. Grady. Random walks for image segmentation. *Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1768–1783, 2006.

[52] D. M. Greig, B. T. Porteous, and A. H. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society*, 51(2):271–279, 1989.

[53] M. Grundmann, V. Kwatra, M. Han, and I. Essa. Efficient hierarchical graph-based video segmentation. In *CVPR'10: Conference on Computer Vision and Pattern Recognition*, pages 1–14, 2010.

[54] G. Guo, N. Zhang, L. Huo, and W. Gao. 2D to 3D conversion based on edge defocus and segmentation. In *ICASSP'08: International Conference on Acoustics, Speech and Signal Processing*, pages 2181–2184, 2008.

[55] Z. Guo, H. Wang, K. Li, Y. Zhang, X. Wang, and Q. Dai. A novel edit propagation algorithm via L0 gradient minimization. In *PCM'15: Pacific-Rim Conference on Advances in Multimedia Information Processing*, pages 402–410, 2015.

[56] M. Guttmann, L. Wolf, and D. Cohen-Or. Semi-automatic stereo extraction from video footage. In *ICCV'09: International Conference on Computer Vision*, pages 136–142, 2009.

[57] K. Han and K. Hong. Geometric and texture cue based depth-map estimation for 2D to 3D image conversion. In *ICCE'11: International Conference on Consumer Electronics*, pages 651–652, 2011.

[58] P. V. Harman, J. Flack, S. Fox, and M. Dowley. Rapid 2D-to-3D conversion. In *EI'02: Electronic Imaging, Stereoscopic Displays and Virtual Reality Systems IX*, pages 78–86, 2002.

[59] K. He, J. Sun, and X. Tang. Guided image filtering. In *ECCV'10: European Conference on Computer Vision*, pages 1–14, 2010.

[60] R. Hebbalaguppe, K. McGuinness, J. Kuklyte, G. Healy, N. O'Connor, and A. Smeaton. How interaction methods affect image segmentation: User experience in the task. In *UCCV'13: Workshop on User-Centred Computer Vision*, pages 19–24, 2013.

[61] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1):185–203, 1981.

[62] A. Hosni, C. Rhemann, M. Bleyer, and M. Gelautz. Temporally consistent disparity and optical flow via efficient spatio-temporal filtering. In *PSIVT'11: Pacific-Rim Symposium on Image and Video Technology*, pages 165–177, 2011.

[63] W. Hu, Z. Dong, and G. D. Yuan. Edit propagation via edge-aware filtering. *Journal of Computer Science and Technology*, 27(4):830–840, 2012.

[64] W. Huang, X. Cao, K. Lu, Q. Dai, and A. C. Bovik. Toward naturalistic 2D-to-3D conversion. *Transactions on Image Processing*, 24(2):724–733, 2015.

[65] I. Ideses, L. Yaroslavsky, and B. Fishbain. Real-time 2D to 3D video conversion. *Journal of Real-Time Image Processing*, 2(1):3–9, 2007.

[66] S. Iizuka, Y. Endo, Y. Kanamori, J. Mitani, and Y. Fukui. Efficient depth propagation for constructing a layered depth image from a single image. *Computer Graphics Forum*, 33(7):279–288, 2014.

[67] M. Ivancics, N. Brosch, and M. Gelautz. Efficient depth propagation in videos with GPU-acceleration. In *VCIP'14: Visual Communications and Image Processing*, pages 1–4, 2014.

[68] M. Ivancsics. Effiziente Tiefenpropagierung in Videos mit GPU-Unterstützung. Master's thesis, Vienna University of Technology, Institute of Software Technology and Interactive Systems, 2014.

[69] L. Jiangbo, S. Keyang, M. Dongbo, L. Liang, and M. N. Do. Cross-based local multipoint filtering. In *CVPR'12: Conference on Computer Vision and Pattern Recognition*, pages 430–437, 2012.

[70] Y. J. Jung, A. Baik, J. Kim, and D. Park. A novel 2D-to-3D conversion technique based on relative height-depth cue. In *EI'09: Electronic Imaging, Stereoscopic Displays and Applications XX*, pages 1–8, 2009.

[71] K. Karsch, C. Liu, and S. Kang. Depth extraction from video using non-parametric sampling. In *ECCV'12: European Conference on Computer Vision*, pages 775–788, 2012.

[72] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

[73] P. Kellnhofer, T. Leimkühler, T. Ritschel, K. Myszkowski, and H. P. Seidel. What makes 2D-to-3D stereo conversion perceptually plausible? In *SAP'15: Symposium on Applied Perception*, pages 59–66, 2015.

[74] J. Kim, A. Baik, Y. J. Jung, and D. Park. 2D-to-3D conversion by using visual attention analysis. In *EI'10: Electronic Imaging, Stereoscopic Displays and Applications XXI*, pages 1–12, 2010.

[75] J. J. Koenderink, A. J. van Doorn, A. M. Kappers, and J. T. Todd. Ambiguity and the 'mental eye' in pictorial relief. *Perception*, 30(4):431–448, 2001.

[76] P. Kohli, H. Nickisch, C. Rother, and C. Rhemann. User-centric learning and evaluation of interactive segmentation systems. *International Journal of Computer Vision*, 100(3):261–274, 2012.

[77] V. Kolmogorov, A. Criminisi, A. Blake, G. Cross, and C. Rother. Bi-layer segmentation of binocular stereo video. In *CVPR'05: Computer Vision and Pattern Recognition*, pages 407–414, 2005.

[78] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele. Joint bilateral upsampling. In *SIGGRAPH'07*, 2007.

[79] V. Kramarev, O. Demetz, C. Schroers, and J. Weickert. Cross anisotropic cost volume filtering for segmentation. In *ACCV'12: Asian Conference on Computer Vision*, pages 803–814, 2013.

[80] L. Lam, S. W. Lee, and C. Y. Suen. Thinning methodologies - a comprehensive survey. *Transactions on Pattern Analysis and Machine Intelligence*, 14(9):869–885, 1992.

[81] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross. Nonlinear disparity mapping for stereoscopic 3D. In *SIGGRAPH'10*, pages 75:1–75:10, 2010.

[82] M. Lang, O. Wang, T. Aydin, A. Smolic, and M. Gross. Practical temporal consistency for image-based graphics applications. *Transactions on Graphics*, 31(4):34:1–34:8, 2012.

[83] S. Y. Lee, J. C. Yoon, and I. K. Lee. Temporally coherent video matting. *Graphical Models*, 72(3):25–33, 2010.

[84] A. Levin, D. Lischinski, and Y. Weiss. Colorization using optimization. In *SIGGRAPH'04*, pages 689–694, 2004.

[85] A. Levin, D. Lischinski, and Y. Weiss. A closed-form solution to natural image matting. *Transactions on Pattern Analysis and Machine Intelligence*, 30(2):228–242, 2008.

[86] J. Lezama, K. Alahari, J. Sivic, and I. Laptev. Track to the future: Spatio-temporal video segmentation with long-range motion cues. In *CVPR'11: Computer Vision and Pattern Recognition*, pages 3369–3376, 2011.

[87] Y. Li, E. Adelson, and A. Agarwala. ScribbleBoost: Adding classification to edge-aware interpolation of local image and video adjustments. In *EGSR'08: Eurographics Conference on Rendering*, pages 1255–1264, 2008.

[88] Y. Li, J. Sun, and H. Y. Shum. Video object cut and paste. In *SIGGRAPH'05*, pages 595–600, 2005.

[89] Y. Li, J. Sun, C. K. Tang, and H. Y. Shum. Lazy snapping. In *SIGGRAPH'04*, pages 303–308, 2004.

[90] Z. Li, X. Cao, and Q. Dai. A novel method for 2D-to-3D video conversion using bi-directional motion estimation. In *ICASSP'12: International Conference on Acoustics, Speech and Signal Processing*, pages 1429–1432, 2012.

[91] Z. Li, X. Xie, and X. Liu. An efficient 2D to 3D video conversion method based on skeleton line tracking. In *3DTV-CON'09: Conference on The True Vision-Capture, Transmission and Display of 3D Video*, pages 1–4, 2009.

[92] M. Liao, J. Gao, R. Yang, and M. Gong. Video Stereolization: Combining motion analysis with user interaction. *Transactions on Visualization and Computer Graphics*, 18(7):1079–1088, 2012.

[93] W. N. Lie, C. Y. Chen, and W. C. Chen. 2D to 3D video conversion with key-frame depth propagation and trilateral filtering. *Electronics Letters*, 47(5):319–321, 2011.

[94] G. Lin, J. Huang, and W. Lie. Semi-automatic 2D-to-3D video conversion based on depth propagation from key-frames. In *ICIP'13: International Conference on Image Processing*, pages 2202–2206, 2013.

[95] D. Lischinski, Z. Farbman, M. Uyttendaele, and R. Szeliski. Interactive local adjustment of tonal values. In *SIGGRAPH'06*, pages 646–653, 2006.

[96] C. Liu. *Beyond pixels: Exploring new representations and applications for motion analysis*. PhD thesis, Electrical Engineering and Computer Science at the Massachusetts Institute of Technology, 2009.

[97] J. Liu, J. Sun, and H. Y. Shum. Paint selection. In *SIGGRAPH'09*, pages 1–7, 2009.

[98] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[99] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence*, pages 674–679, 1981.

[100] L. Q. Ma and K. Xu. Efficient manifold preserving edit propagation with adaptive neighborhood size. *Computers and Graphics*, 38:167–173, 2014.

[101] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Henry Holt and Co., Inc., 1982.

[102] S. Martull, M. P. Martorell, and K. Fukui. Realistic CG stereo image dataset with ground truth disparity maps. In *ICPR20'12: International Conference on Pattern Recognition*, pages 1038–1042, 2012.

[103] Microsoft Kinect. http://www.microsoft.com/en-us/kinectforwindows/. Accessed: 2014-05-28.

[104] E. N. Mortensen and W. A. Barrett. Intelligent scissors for image composition. In *SIGGRAPH'95*, pages 191–198, 1995.

[105] K. Moustakas, D. Tzovaras, and M. G. Strintzis. Stereoscopic video generation based on efficient layered structure and motion estimation from a monoscopic image sequence. *Transactions on Circuits and Systems for Video Technology*, 15:1065–1073, 2005.

[106] M. Nezveda, N. Brosch, F. Seitner, and M. Gelautz. Depth map post-processing for depth-image-based rendering: A user study. In *EI'14: Electronic Imaging, Stereoscopic Displays and Applications XXV*, pages 1–9, 2014.

[107] G. Noris, D. Sykora, A. Shamir, S. Coros, B. Whited, M. Simmons, A. Hornung, M. Gross, and R. Sumner. Smart scribbles for sketch segmentation. *Computer Graphics Forum*, 31(8):2516–2527, 2012.

[108] Nvidia. CUDA: Compute unified device architecture programming guide. Technical report, 2008.

[109] P. Ochs and T. Brox. Object segmentation in video: A hierarchical variational approach for turning point trajectories into dense regions. In *ICCV'11: International Conference on Computer Vision*, pages 1583–1590, 2011.

[110] A. S. Ogale and Y. Aloimonos. A roadmap to the integration of early visual modules. *International Journal of Computer Vision*, 72:9–25, 2007.

[111] T. Okino, H. Murata, K. Taima, T. Iinuma, and K. Oketani. New television with 2D/3D image conversion technologies. In *EI'96: Electronic Imaging, Stereoscopic Displays and Virtual Reality Systems*, pages 96–103, 1996.

[112] G. Palou and P. Salembier. Depth order estimation for video frames using motion occlusions. *Computer Vision*, 8(2):152–160, 2013.

[113] S. Paris. Edge-preserving smoothing and mean-shift segmentation of video streams. In *ECCV'08: European Conference on Computer Vision: Part II*, pages 460–473, 2008.

[114] S. Paris and F. Durand. A topological approach to hierarchical segmentation using mean shift. In *CVPR'07: Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[115] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama. Digital photography with flash and no-flash image pairs. In *SIGGRAPH'04*, volume 23, pages 664–672, 2004.

[116] R. Phan and D. Androutsos. A semi-automatic 2D to stereoscopic 3D image and video conversion system in a semi-automated segmentation perspective. In *EI'13: Electronic Imaging, Stereoscopic Displays and Applications XXIV*, pages 1–12, 2013.

[117] R. Phan and D. Androutsos. Robust semi-automatic depth map generation in unconstrained images and video sequences for 2D to stereoscopic 3D conversion. *Transactions on Multimedia*, 16(1):122–136, 2014.

[118] R. Phan, R. Rzeszutek, and D. Androutsos. Semi-automatic 2D to 3D image conversion using scale-space random walks and a graph cuts based depth prior. In *ICIP'11: International Conference on Image Processing*, pages 865–868, 2011.

[119] R. Phan, R. Rzeszutek, and D. Androutsos. *Multimedia Image and Video Processing*, chapter Literature survey on recent methods for 2D to 3D video conversion, pages 691–716. CRC Press, 2nd edition, 2012.

[120] J. C. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3):61–74, 1999.

[121] R. Poppe. A survey on vision-based human action recognition. *Image and Vision Computing*, 28(6):976–990, 2010.

[122] B. L. Price, B. S. Morse, and S. Cohen. LIVEcut: Learning-based interactive video segmentation by evaluation of multiple propagated cues. In *ICCV'09: International Conference on Computer Vision*, pages 779–786, 2009.

[123] S. J. D. Prince. *Computer Vision: Models, Learning, and Inference*. Cambridge University Press, 1st edition, 2012.

[124] F. Remondino and D. Stoppa. *ToF range-imaging cameras*. Springer, 2012.

[125] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz. Fast cost-volume filtering for visual correspondence and beyond. In *CVPR'11: Conference on Computer Vision and Pattern Recognition*, pages 3017–3024, 2011.

[126] C. Richardt, D. Orr, I. Davies, A. Criminisi, and N. A. Dodgson. Real-time spatiotemporal stereo matching using the dual-cross-bilateral grid. In *ECCV'10: European Conference on Computer Vision*, pages 510–523, 2010.

[127] C. Rother, V. Kolmogorov, and A. Blake. GrabCut: Interactive foreground extraction using iterated graph cuts. In *SIGGRAPH'04*, pages 309–314, 2004.

[128] C. Rusu and S. A. Tsaftaris. Estimation of scribble placement for painting colorization. In *ISPA'13: International Symposium on Image and Signal Processing and Analysis*, pages 564–569, 2013.

[129] R. Rzeszutek and D. Androutsos. Label propagation through edge-preserving filters. In *ICASSP'14: International Conference on Acoustics, Speech and Signal Processing*, pages 599–603, 2014.

[130] R. Rzeszutek, R. Phan, and D. Androutsos. Semi-automatic synthetic depth map generation for video using random walks. In *ICME'11: International Conference on Multimedia and Expo*, pages 1–6, 2011.

[131] P. Sand and S. Teller. Particle video: Long-range motion estimation using point trajectories. In *CVPR'06: Conference on Computer Vision and Pattern Recognition*, pages 2195–2202, 2006.

[132] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nešić, X. Wang, and P. Westling. High-resolution stereo datasets with subpixel-accurate ground truth. In *GCPR'14: German Conference Pattern Recognition*, pages 31–42, 2014.

[133] D. Scharstein and C. Pal. Learning conditional random fields for stereo. In *CVPR'07: Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[134] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47:7–42, 2002.

[135] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *CVPR'03: Computer Vision and Pattern Recognition*, pages 195–202, 2003.

[136] T. Schausberger. Temporally coherent cost volume filtering-based depth propagation in videos. Master's thesis, Vienna University of Technology, Institute of Software Technology and Interactive Systems, 2015.

[137] M. Seymour. Art of stereo conversion: 2D to 3D - 2012. https://www.fxguide.com/featured/art-of-stereo-conversion-2d-to-3d-2012/. Accessed: 2014-04-29.

[138] E. Sharon, M. Galun, D. Sharon, R. Basri, and A. Brandt. Hierarchy and adaptivity in segmenting visual scenes. *Nature*, 442(7104):810–813, 2006.

[139] J. Shi and J. Malik. Normalized cuts and image segmentation. *Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[140] Sky 3D. http://www.sky.at/3D. Accessed: 2014-04-25.

[141] A. Smolic, P. Kauff, S. Knorr, A. Hornung, M. Kunter, M. Muller, and M. Lang. Three-dimensional video postproduction and processing. *Proceedings of the IEEE*, 99(4):607–625, 2011.

[142] A. N. Stein, T. S. Stepleton, and M. Hebert. Towards unsupervised whole-object segmentation: Combining automated matting with boundary detection. In *CVPR'08: Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[143] StereoD. http://www.stereodllc.com/. Accessed: 2014-04-29.

[144] Stereoscopic Suite X. http://www.emotion3d.tv Accessed: 2014-08-12.

[145] N. Sundaram, T. Brox, and K. Keutzer. Dense point trajectories by GPU-accelerated large displacement optical flow. In *ECCV'10: European Conference on Computer Vision: Part I*, pages 438–451, 2010.

[146] R. Szeliski. *Computer Vision: Algorithms and Applications*. Springer-Verlag New York, 1st edition, 2010.

[147] W. J. Tam, C. Vázquez, and F. Speranza. Three-dimensional TV: A novel method for generating surrogate depth maps using color information. In *EI'09: Electronic Imaging, Stereoscopic Displays and Applications XX*, pages 1–9, 2009.

[148] R. Tarjan. Depth-first search and linear graph algorithms. In *Journal on Computing*, volume 1, pages 146–160, 1972.

[149] D. Terzopoulos and R. Szeliski. Tracking with kalman snakes. In *Active Vision*, pages 3–20. MIT Press, 1993.

[150] A. N. Tikhonov and V. Y. Arsenin. *Solutions of ill-posed problems*. V. H. Winston & Sons, 1977.

[151] D. A. Tolliver and G. L. Miller. Graph partitioning by spectral rounding: Applications in image segmentation and clustering. In *CVPR'06: Computer Vision and Pattern Recognition*, pages 1053–1060, 2006.

[152] C. Tomasi and T. Kanade. *Detection and Tracking of Point Features*. Shape and motion from image streams. School of Computer Science, Carnegie Mellon Univ., 1991.

[153] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *ICCV'98: International Conference on Computer Vision*, pages 839–846, 1998.

[154] E. Turetken and A. Alatan. Temporally consistent layer depth ordering via pixel voting for pseudo 3D representation. In *3DTV-CON'09: 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video*, pages 1–4, 2009.

[155] A. van Doorn, J. Koenderink, and J. Wagemans. Rank order scaling of pictorial depth. *Iperception*, 2(7):724–44, 2011.

[156] C. Varekamp and B. Barenbrug. Improved depth propagation for 2D to 3D video conversion using key-frames. In *IETCVMP'07: European Conference on Visual Media Production*, pages 1–7, 2007.

[157] S. Vicente, V. Kolmogorov, and C. Rother. Graph cut based image segmentation with connectivity priors. In *CVPR'08: Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[158] A. Voronov, D. Vatolin, D. Sumin, V. Napadovsky, and A. Borisov. Methodology for stereoscopic motion-picture quality assessment. In *EI'13: Electronic Imaging, Stereoscopic Displays and Applications XXIV*, pages 1–14, 2013.

[159] D. Wang, J. Liu, Y. Ren, C. Ge, W. Liu, and Y. Li. Depth propagation based on depth consistency. In *WCSP'12: International Conference on Wireless Communications Signal Processing*, pages 1–6, 2012.

[160] J. Wang, P. Bhat, R. A. Colburn, M. Agrawala, and M. F. Cohen. Interactive video cutout. In *SIGGRAPH'05*, pages 585–594, 2005.

[161] J. Wang, B. Thiesson, Y. Xu, and M. Cohen. Image and video segmentation by anisotropic kernel mean shift. In *ECCV'04: European Conference on Computer Vision*, pages 238–249, 2004.

[162] J. Wang, Y. Xu, and M. F. Shum, H. Y. Cohen. Video tooning. In *SIGGRAPH'04*, pages 574–583, 2004.

[163] O. Wang, M. Lang, M. Frei, A. Hornung, A. Smolic, and M. Gross. StereoBrush: Interactive 2D to 3D conversion using discontinuous warps. In *SBIM'11: Eurographics Symposium on Sketch-Based Interfaces and Modeling*, pages 47–54, 2011.

[164] B. Ward, S. Kang, and E. P. Bennett. Depth Director: A System for Adding Depth to Movies. *Computer Graphics and Applications*, 31(1):36–48, 2011.

[165] C. Wu, G. Er, X. Xie, T. Li, X. Cao, and Q. Dai. A novel method for semi-automatic 2D to 3D video conversion. In *3DTV-CON'08: Conference on The True Vision-Capture, Transmission and Display of 3D Video*, pages 65–68, 2008.

[166] J. Wulff, D. J. Butler, G. B. Stanley, and M. J. Black. Lessons and insights from creating a synthetic optical flow benchmark. In *ECCV'12: European Conference on Computer Vision*, pages 168–177, 2012.

[167] L. Xu, Q. Yan, and J. Jia. A sparse control model for image and video editing. *Transactions on Graphics*, 32(6):197:1–197:10, 2013.

[168] P. Xu, H. Fu, O. K. C. Au, and C. L. Tai. Lazy Selection: A scribble-based tool for smart shape elements selection. *Transactions on Graphics*, 31(6):142:1–142:9, 2012.

[169] X. Xu, L. Po, K. Cheung, K. Ng, K. Wong, and C. Ting. Watershed and random walks based depth estimation for semi-automatic 2D to 3D image conversion. In *ICSPCC'12: International Conference on Signal Processing, Communication and Computing*, pages 84–87, 2012.

[170] L. Yatziv and G. Sapiro. Fast image and video colorization using chrominance blending. *Transactions on Image Processing*, 15(5):1120–1129, 2006.

[171] A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. *Journal of Computing Surveys*, 38, 2006.

[172] YouTube, LLC. https://www.youtube.com. Accessed: 2014-04-25.

[173] J. Zhang and Z. Zhang. Depth map propagation with the texture image guidance. In *ICIP'14: International Conference on Image Processing*, pages 3813–3817, 2014.

[174] L. Zhang, C. Vázquez, and S. Knorr. 3D-TV content creation: Automatic 2D-to-3D video conversion. *Transactions on Broadcasting*, 57(2):372–383, 2011.

[175] Z. Zhang, C. Zhou, B. Xin, Y. Wang, and W. Gao. An interactive system of stereoscopic video conversion. In *MM'12: International Conference on Multimedia*, pages 149–158, 2012.

[176] Avatar. Dir. J. Cameron, *20th Century Fox*. Film, 2009.

[177] Toy Story 2. Dir. J. Lasseter, *Disney Digital 3-D*. Film, 2010.

[178] Star Wars: Episode I. Dir. G. Lucas, *20th Century Fox*. Film, 2012.