B

Binocular Stereo

Tatsunori Taniai
Research Administrative Division, OMRON SINIC X Corporation, Tokyo, Japan
Discrete Optimization Unit, RIKEN Center for Advanced Intelligence Project, Tokyo, Japan

Synonyms

Stereo Matching, Binocular Stereo Vision

Related Concepts

▶Multiview Stereo; ▶Wide Baseline Matching; ▶Dense Reconstruction; ▶Epipolar Geometry ▶Photo-Consistency; ▶Occlusion Handling; ▶Subpixel Estimation;

Definition

Binocular stereo refers to the task of recovering depths of a static scene using a pair of overlapping images captured from different viewpoints. Binocular stereo systems usually use two identical parallel cameras that are horizontally separated by a certain distance, referred to as the *baseline*. The task of binocular stereo amounts to finding dense pixel correspondences between the image pair along horizontal scan-lines (called *epipolar lines*) or estimating the *disparity* for each pixel of the stereo images. The outcome of binocular stereo takes a form of a depth map that can be computed from disparity given the baseline and focal length of a stereo system or instead a disparity map itself.

Background

Binocular stereo is one of the oldest topics in computer vision. Similar to the mechanism of human depth perception, the principle of binocular stereo is triangulation, which is mathematically formulated based on the epipolar geometry. However, due to ambiguous pixel correspondences, binocular stereo typically becomes ill-posed when scenes have no textures or show repetitive patterns.

Wide baseline stereo refers to a particular setting of binocular stereo where the two cameras are widely separated. This leads to a more difficult task because of larger disparity ranges, lesser overlaps between image pairs, and a higher likelihood of occlusions.

Occlusion is an inevitable problem in binocular stereo, which occurs when a part of an object in one view is not present in the other view because it is occluded by another object or is out of the field-of-view in the other view. Since occluded regions have no true visual correspondence, they produce incorrect depth estimates unless properly handled.

Binocular stereo can be seen as special or simplified cases of other related computer vision tasks. For example, multiview stereo uses two or more images captured from calibrated but possibly irregular viewpoints. In principle, multiview stereo can achieve higher accuracy than binocular stereo,

K. Ikeuchi (ed.), *Computer Vision*, DOI 10.1007/978-0-387-31439-6,

because the use of multiple images can reduce matching ambiguities and can also lead to fewer occluded surfaces (as each surface point has a better chance of being visible from at least two views). However, multiview stereo is a more complicated task, because images from irregular viewpoints may contain low-overlapping image pairs that have to be excluded from matching via viewpoint selection. Also, surface patches often undergo more significant distortions across views, which makes accurate evaluations of patch similarity difficult.

Optical flow is also a visual correspondence estimation task between two images, but involves estimating more general motions of a dynamic scene between two different temporal-frame images captured by a possibly-moving monocular camera. While pixel motions in binocular stereo (disparities) are induced by factors of object positions and the left-to-right camera motion, optical flow involves more complicated motion factors of object positions, an unknown camera motion, and dynamic object movements. Because of this complexity, estimation of pixel motions in optical flow requires a 2D search space, which is wider than 1D search spaces for disparities and depths in binocular and multiview stereo. The presence of dynamic object movements also makes the occlusion reasoning more complicated than stereo.

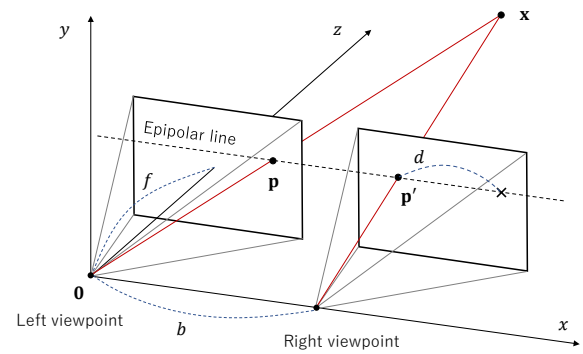
Binocular stereo can thus be considered as the most fundamental dense visual correspondence estimation task, which is built upon notions from a wide range of computer vision areas such as camera calibration, image filtering, and combinatorial optimization as explained in following sections.

Theory and Application

We first discuss the mathematical foundation of binocular stereo based on the epipolar geometry. We then review classical methodologies of binocular stereo, and also review recent learning based methodologies using neural networks.

a) Mathematical Principle

In this section, we explain how a 3D point can be triangulated from a pair of corresponding pixels, using a typical *rectified* setting of binocular stereo shown in Fig. 1. Here, two identical pinhole cameras, both directed along the z axis in the 3D world coordinate system, are positioned at the origin $(0, 0, 0)^T$ (left viewpoint) and a horizontally shifted place $(b, 0, 0)^T$ (right viewpoint) where b is baseline. Both cameras are calibrated and have the following intrinsic parameter matrix



Binocular Stereo, Fig. 1 Rectified setting of binocular stereo where two parallel cameras are horizontally placed.

$$K = \begin{bmatrix} f & 0 & c_u \\ 0 & f & c_v \\ 0 & 0 & 1 \end{bmatrix}, \tag{1}$$

where (c_u, c_v) is the principal point and f is the focal length. Note that in reality it is difficult to setup such an ideally rectified stereo capturing system. However, given a calibrated stereo system (*i.e.*, the relative pose and intrinsic parameters of the two cameras are known), we can transform unrectified stereo image pairs into rectified ones using a technique of stereo image rectification.

In this rectified setting, suppose there is a surface shown at a pixel $\mathbf{p} = (u, v)^T$ in the left view image (reference image). Its unknown 3D coordinate $\mathbf{x} = (x, y, z)^T$ in question can be represented as

$$\mathbf{x} = K^{-1}(z\bar{\mathbf{p}}), \tag{2}$$

where $\bar{\mathbf{p}}$ is the homogeneous coordinate of $\mathbf{p}$. This 3D point $\mathbf{x}$ can be projected to the right view image at the following 2D location $\mathbf{p}' = (u', v')^T$.

$$\mathbf{p}' = \pi(K \begin{bmatrix} R & \mathbf{t} \end{bmatrix} \bar{\mathbf{x}}). \tag{3}$$

Here, R is the identity rotation matrix, $\mathbf{t} = (-b, 0, 0)^T$ is the translation representing horizontal baseline, and π is a function $\pi(x, y, z) = (x/z, y/z)^T$. Thus, by plugging Eq. 2 into Eq. 3, we obtain an expression for the point corresponding to $\mathbf{p}$ in the right image as

$$\mathbf{p}' = \mathbf{p} - \begin{bmatrix} fb/z \\ 0 \end{bmatrix}. \tag{4}$$

This result shows that for each pixel $\mathbf{p}$ in the left image, its correspondence $\mathbf{p}'$ in the right image should be found at a horizontally-shifted position of $\mathbf{p}$ by shifting the pixel by the following amount to the left.

$$d = fb/z. \tag{5}$$

This horizontal shifting amount d is called *disparity*. As shown by Eq. 5, once we obtain a disparity d for a pixel (or obtain its correspondence $\mathbf{p}'$), we can obtain its depth z from the disparity given the baseline b and focal length f of the considered stereo system.

b) Classical Methodologies

Scharstein and Szeliski [11] provide a well known taxonomy of classical stereo algorithms based on the following four steps of algorithms: matching cost computation (photo-consistency), cost aggregation, disparity computation and optimization, and disparity refinement. In this section, we discuss classical stereo algorithms in terms of design and minimization of the following objective function

$$E(\mathbf{D}) = \sum_{\mathbf{p}} C_{\mathbf{p}}(D_{\mathbf{p}}) + R(\mathbf{D}). \tag{6}$$

Here, $\mathbf{D}$ represents a disparity map that we estimate for an input stereo image pair. $C_{\mathbf{p}}(D_{\mathbf{p}})$ is a matching-cost term that evaluates a given disparity estimate $D_{\mathbf{p}}$ for a pixel $\mathbf{p}$ by measuring photo-consistencies between the two images. $R(\mathbf{D})$ is a regularization term that enforces some notion of smoothness on the disparity map $\mathbf{D}$.

The disparity map $\mathbf{D}$ often takes a discrete variable form $\mathbf{D} \in \{d_1, d_2, ..., d_K\}^{H \times W}$, and thus the objective function $E(\mathbf{D})$ is optimized using discrete (combinatorial) optimization algorithms. This is because objective functions of stereo are usually highly non-convex and continuous optimization methods can be easily trapped at poor local minima.

Stereo algorithms can be broadly divided into local and global methods. Local methods rely only on the matching-cost term and minimize the objective function by a simple winner-takes-all strategy. Global methods use a more complicated objective function with explicit regularizers, which involve computationally more expensive optimization procedures. Below we review important components and techniques of those local and global stereo methods.

Photo-consistency. As a critical component of the matching-cost term, photo-consistency $\rho(\mathbf{p}, \mathbf{p}')$ is a scalar function that evaluates dissimilarity between two pixels or image patches at respective locations $\mathbf{p}$ and $\mathbf{p}'$ in a given image pair $\{\mathbf{I}, \mathbf{I}'\}$. The simplest photo-consistency measure is SAD (sum of absolute difference) that evaluates $\rho(\mathbf{p}, \mathbf{p}') = |I_{\mathbf{p}} - I'_{\mathbf{p}'}|$, but directly comparing image intensities is not robust to illumination changes. As a more robust measure, normalized cross correlation (NCC) is used to compare two image patches. Zabih and Woodfill [18] propose the CENSUS transform that encodes an image patch into a binary feature vector, whose dissimilarity can be efficiently computed as the Hamming distance.

Cost aggregation. Cost aggregation refers to a technique to refine noisy raw photo-consistency measures $\rho(\mathbf{p}, \mathbf{p}')$ by summing them over pixels in a patch around $\mathbf{p}$ as

$$C_{\mathbf{p}}(d) = \sum_{\mathbf{s} \in W_{\mathbf{p}}} \omega_{\mathbf{ps}} \rho(\mathbf{s}, \mathbf{s}'_d). \tag{7}$$

Here, $W_{\mathbf{p}}$ is a support window centered at $\mathbf{p}$ in the reference image, $\omega_{\mathbf{ps}}$ is some weight function, and $\mathbf{s}'_d = \mathbf{s} - (d, 0)^T$. Cost aggregation is often referred to as *cost volume filtering*, because if we precompute raw matching costs $\rho(\mathbf{p}, \mathbf{p}'_d)$ for all pixels $\mathbf{p}$ and for all pre-defined disparities $d \in \{d_1, d_2, ..., d_K\}$ as a 3D cost volume $V(\mathbf{p}, d)$, then cost aggregation is carried out by applying an image filter on 2D cost map slices $V_d(\mathbf{p}) = V(\mathbf{p}, d)$ with a filter kernel of $\omega_{\mathbf{ps}}$. Although a naive implementation of cost aggregation requires $O(|W_{\mathbf{p}}|)$ of computations for each term $C_{\mathbf{p}}(d)$, the notion of

(a) Support window (b) Depth (e) Bilateral weights (c) Spatial weights (d) Color-based weights

Binocular Stereo, Fig. 2 Adaptive support windows. For (a) an support window with (b) depth boundaries, the adaptive window method [17] computes (e) bilateral support weights by combining (c) spatial weights and (d) color-based weights.

cost volume filtering can allow $O(1)$ of computations when using a constant-time filter (*e.g.*, a box filter $\omega_{\mathbf{ps}} = 1$).

Cost aggregation relies on an assumption that the support pixels $\mathbf{s}$ in a window $W_{\mathbf{p}}$ have the same disparity. However, as discussed in [3], this assumption often breaks down in two cases: 1) when there are depth boundaries in the window; 2) when the window region shows a highly slanted surface that has significantly varying disparities.

The first issue causes boundary-flattening artifacts in resulting disparity maps, but it can be well handled by adaptive window approaches [17] that use soft support window weights $\omega_{\mathbf{ps}}$ for cost aggregation. Yoon and Kweon [17] propose to use the joint bilateral filtering for cost aggregation as illustrated in Fig. 2.

The second issue causes staircase artifacts at slated surfaces especially when large support windows are used. For this, Bleyer *et al.* [3] propose a slanted patch-matching technique, which approximates a surface in a support window by linearly-varying disparities (parameterized by a disparity plane $d = au + bv + c$) instead a constant disparity and can thus relax the fronto-parallel window bias. This approach imposes a complicated inference task of pixelwise 3D continuous variables (a, b, c), which is solved by inference techniques explained later in a section on continuous disparity estimation.

Regularization. Local methods that only rely on the matching-cost term often produce inaccurate disparities due to low feature regions or noises of matching costs. Therefore, global methods add a regularization term $R(\mathbf{D})$ in the objective function that is minimized by an optimization algorithm.

A widely adopted regularization is the truncated linear model

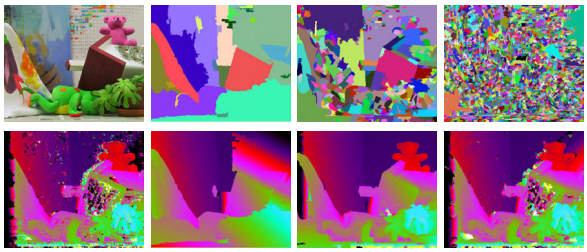
$$R(\mathbf{D}) = \sum_{(\mathbf{p},\mathbf{q})\in N} \omega_{\mathbf{pq}} \min\{\tau, |D_{\mathbf{p}} - D_{\mathbf{q}}|\}, \qquad (8)$$

where N is the set of neighboring pixel pairs, $\omega_{\mathbf{pq}}$ is a contrast-sensitive weight for preserving edges, and τ is a user-defined threshold parameter for allowing depth jumps at object boundaries. Because of its simple pairwise function form, adopting this model can keep the optimization quite tractable [13]. However, it is known to have the fronto-parallel bias causing staircase artifacts at slanted surfaces [16].

A variety of regularizers have been proposed to handle the fronto-parallel bias. Woodford *et al.* [16] propose a second-order smoothness term, which evaluates $|D_{\mathbf{q}} - 2D_{\mathbf{p}} + D_{\mathbf{r}}|$ instead of $|D_{\mathbf{p}} - D_{\mathbf{q}}|$ for three consecutive pixels $(\mathbf{q}, \mathbf{p}, \mathbf{r})$. However, it imposes complicated optimization due to the higher-order form of the objective function and treatment of continuous disparities. Olsson *et al.* [10] propose a powerful curvature regularization term, which requires pixelwise continuous disparity plane estimation but allows an efficient pairwise function form. Scharstein *et al.* [12] propose a scheme to encode pre-estimated surface orientation priors into regularization without increasing computational costs of optimization.

Optimization. Optimization is a necessary step in global methods to minimize their objective function with pairwise or higher-order interaction terms for enforcing regularization.

When disparities are discrete variables and the objective function has only up to their pairwise



Binocular Stereo, Fig. 3 Segment-based disparity map proposals for fusion. The image courtesy of Woodford *et al.* [16].

interactions, then its optimization is well established [13]; we can directly apply discrete optimizers such as message passing algorithms (belief propagation) and the expansion move algorithm using graph cuts to approximately solve the combinatorial optimization problem. A commonly used practical optimization method is the semi-global matching (SGM) [5], which has a good trade-off property between accuracy and efficiency for real-time applications. It has been shown to be a variant of message-passing techniques [4].

Continuous disparity estimation. Because disparities inherently reside in a continuous space, we need to infer continuous disparities for a more accurate representation of 3D scenes.

One type of approaches to continuous disparity estimation seeks continuous disparities during optimization. Since discrete optimizers cannot be directly applied for this purpose, discrete-continuous optimization strategies are often employed. For example, segment-based stereo [2] optimizes the assignment of pre-estimated disparity plane labels to each of superpixel regions[1], which produces continuous-valued but piecewise planar disparity maps. Fusion based methods [16] fuse many continuous-valued disparity map proposals to produce a better solution by solving a combinatorial optimization task using graph cuts, where proposals are generated, *e.g.*, by segment-based methods using various patterns of superpixels (see Fig. 3 for an illustration). PatchMatch stereo [3] estimates pixelwise continuous disparity planes using a randomized search scheme, which no longer requires pre-estimated proposals. Its variants using belief propagation [1] or graph cuts [14] further add regularization into this randomized search scheme.

Another type of approaches estimates continuous disparities as post-processing by refining initial discrete disparity estimates. Since it is usually used to refine initial disparities at integer pixels, this refinement process is often called *subpixel refinement*. For example, techniques based on gradient descent [8] or curve (parabola) fitting [5] are often employed.

Occlusion handling. Occlusion handling in binocular stereo is usually done either as post processing using left right consistency check [3, 5] or during optimization by incorporating a occlusion model into the objective function [7, 15]. While the former approach can be adopted for both local and global methods, the latter can be only employed by global methods at the cost of producing complicated higher-order interactions in the matching-cost term.

c) Learning based Methodologies

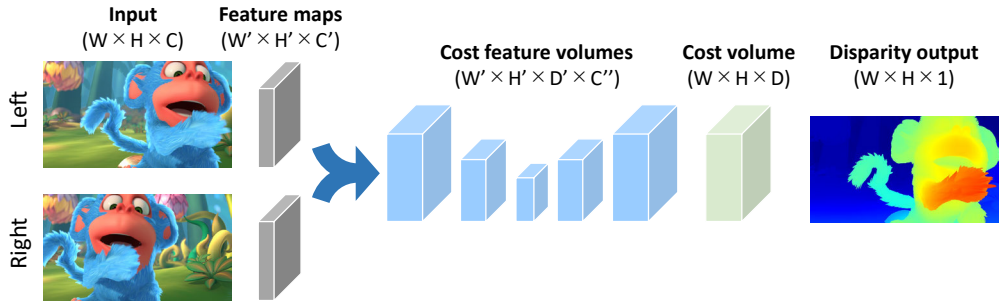
As an emerging trend in this field, learning based approaches to binocular stereo have been gathering great interest. In particular, end-to-end learning approaches using deep neural networks are popular, which directly learn a mapping function f from an image pair to a disparity map as

$$\mathbf{D} = f(\mathbf{I}, \mathbf{I}'; \boldsymbol{\Theta}). \tag{9}$$

The function f is implemented as convolutional neural networks (CNN), whose parameters $\boldsymbol{\Theta}$ are optimized so as to minimize some loss function $\ell(\mathbf{D})$ over a large amount of training data. The loss is evaluated using ground truth disparities in supervised learning, or using a criteria similar to the classical objective function in Eq. 6 without using ground truth in self-supervised learning.

Such learning based methods are often advantageous over classical methods in that they can automatically handle difficulties of stereo such as image patch distortions, illumination changes, occlusions, rectification and calibration errors, by

[1] In segment-based methods, the objective function in Eq. 6 is modified so that each node $\mathbf{p}$ and variable $\mathbf{D}_\mathbf{p}$ represents a superpixel and its disparity plane assignment, respectively.



Binocular Stereo, Fig. 4 Basic neural network architecture for end-to-end learning of binocular stereo.

data-driven approaches. Because computations of CNNs are massively parallelizable on GPUs, CNN based methods can perform quite efficiently even for continuous disparity inference.

Analogously to the taxonomy of classical stereo algorithms [11], we identify four stages that neural network architectures for binocular stereo often perform: feature extraction, volume construction, cost volume learning, and disparity computation and refinement. An example architecture of such neural networks is shown in Fig. 4. Based on this view, below we review existing works on learning based methods.

Feature extraction. Early works on learning based methods use a neural network to compute stereo matching costs. MC-CNN by Zbontar and LeCun [19] extracts feature vectors from image patches and computes matching costs using the cosine distance (fast version) or fully-connected layers (accurate version). Learned matching costs are then fed into classical stereo pipelines (SGM [5]) for disparity estimation. Later, this feature extraction is taken as the first stage of network architectures in end-to-end learning approaches [6, 9, 20], *e.g.*, using feed-forward CNNs [9], ResNet-like networks [6], spatial pyramid pooling (SSP) layers, or 2D hourglass networks [20], with the following subsequent stages.

Volume construction. A seminal work by Mayer *et al.* [9] proposes DispNetC, which constructs a matching cost volume and regresses out a continuous disparity map for end-to-end learning. Volume construction is initially done in [9] by correlating left and traversed right feature maps, but it is later extended to concatenate feature maps [6] or combine concatenation and group-wise correlation.

Cost volume learning. Kendall *et al.* [6] propose GC-Net, which processes a concatenation based cost feature volume (4D tensor) by a 3D hourglass network using 3D Conv layers for cost volume learning. Zhang *et al.* [20] propopse semi-global aggregation and local guided aggregation layers for cost volume learning, analogously to classical techniques of SGM [5] and adaptive window based cost aggregation [17].

Disparity computation and refinement. In early works [9, 19], computing disparities is not explicitly done in classification based methods [19] or done by treating a 3D cost volume as a 2D feature map for a scalar-map regression CNN in DispNetC [9]. Kendall *et al.* [6] propose the soft-argmin operator that can more effectively output a continuous disparity map from a 3D cost volume. Regressed disparity maps are often further processed by a shallow 2D CNN for refinement.

Open Problems

Benchmarks and Datasets

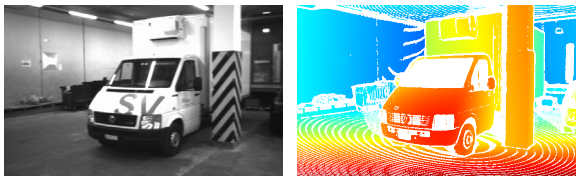
As there is an increasing demand for large-scale binocular stereo datasets for data-driven approaches, we introduce benchmarks and datasets that are popularly used in this area.

Middlebury benchmark (version 3)[2] provides high resolution (1 to 6 MPix) stereo image pairs

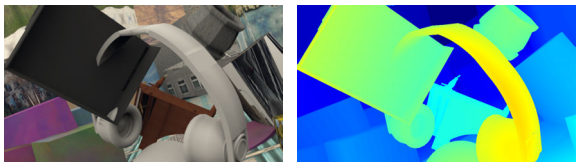
[2] http://vision.middlebury.edu/stereo/

Binocular Stereo, Fig. 5 Middlebury benchmark (an example scene image and its ground truth disparity map).

Binocular Stereo, Fig. 6 KITTI 2015 and 2012 benchmarks.



Binocular Stereo, Fig. 7 ETH3D benchmark.



Binocular Stereo, Fig. 8 SceneFlow dataset.

of 10 testing and 23 training indoor scenes with highly accurate dense ground truth disparities obtained by using structured-light scanning. An example is shown in Fig. 5. The benchmark is designed to contain some challenges such as different exposures or illumination conditions between image pairs, and the presence of vertical displacements due to imperfect rectification.

KITTI 2015 benchmark[3] provides 200 stereo image pairs of 376 × 1242 pixels (0.5 MPix) for each of training and testing sets, recorded by a synchronized stereo camera mounted on a vehicle running on public roads. An example is shown in Fig. 6 (top). The training set images are provided with ground truth disparities for background (using sparse depths measured by a LiDAR sensor) and foreground regions (using 3D CAD models of vehicles manually registered to the scenes).

KITTI 2012 benchmark[4] provides stereo image pairs of 194 training and 195 testing scenes. The images and ground truth disparities are provided similarly to the 2015 version. An example is shown in Fig. 6 (bottom).

ETH3D benchmark[5] (low-res two-view) provides 27 training and 20 testing stereo image pairs, covering both indoor and outdoor scenes. Each image has about 0.4 MPix. Training images are provided with ground truth disparities obtained by a laser scanner. An example is shown in Fig. 7.

SceneFlow dataset[6] is a synthetic large scale dataset for neural network training. It provides stereo image pairs of totally 35,454 training and 4,370 testing scenes, where each image has 960 × 540 pixels (0.5 MPix) and is provided with the ground truth disparities for all the pixels. Both sets contain randomly generated 3D scenes named *FlyingThings3D* (see Fig. 8 for an example) and the training set further contains additional *Monkaa* (see Fig. 4) and *Driving* subsets.

References

1. Besse, F., Rother, C., Fitzgibbon, A.W., and Kautz, J. (2014). PMBP: PatchMatch belief propagation for correspondence field estimation. Int'l J. Comp. Vis. (IJCV) 110(1), 2–13.
2. Birchfield, S. and Tomasi, C. (1999). Multiway cut for stereo and motion with slanted surfaces. In: Proc. Int'l Conf. Comp. Vis. (ICCV). pp. 489–495.
3. Bleyer, M., Rhemann, C., abd Rother, C. (2011). PatchMatch stereo - Stereo matching with slanted support windows. In: Proc. British Mach. Vis. Conf. (BMVC). pp. 1–11.

[3] http://www.cvlibs.net/datasets/kitti/eval_scene_flow.php?benchmark=stereo

[4] http://www.cvlibs.net/datasets/kitti/eval_stereo_flow.php?benchmark=stereo

[5] https://www.eth3d.net

[6] https://lmb.informatik.uni-freiburg.de/resources/datasets/SceneFlowDatasets.en.html

4. Drory, A., Haubold, C., Avidan, S., and Hamprecht, F. A. (2014). Semi-global matching: a principled derivation in terms of message passing. Pattern Recognition, 43–53.
5. Hirschmuller, H. (2008). Stereo processing by semiglobal matching and mutual information. IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI) 30(2), 328–341.
6. Kendall, A., Martirosyan, H., Dasgupta, S., Henry, P., Kennedy, R., Bachrach, A., and Bry, A. (2017). End-to-end learning of geometry and context for deep stereo regression. In: Proc. Int'l Conf. Comp. Vis. (ICCV).
7. Kolmogorov, V. and Zabih, R. (2001). Computing visual correspondence with occlusions using graph cuts. In: Proc. Int. Conf. Comp. Vis. (ICCV), volume 2, 508–515.
8. Lucas, B.D. and Kanade, T. (1981). An iterative image registration technique with an ap-plication to stereo vision. In: Hayes, P.J. (ed.) Proc. of Int'l Joint Conf. Art. Intell. (IJCAI). pp. 674–679.
9. Mayer, N., Ilg, E., Hausser, P., Fischer, P., Cremers, D., Dosovitskiy, A., and Brox, T. (2016). A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation. In: Proc. IEEE Conf. Comp. Vis. Pattern Recog. (CVPR).
10. Olsson, C., Ulen, J., and Boykov, Y. (2013). In Defense of 3D-Label Stereo. In: Proc. IEEE Conf. Comp. Vis. Pattern Recog. (CVPR). 1730–1737.
11. Scharstein, D. and Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Comp. Vis. (IJCV) **47** (1/2/3), 7–42.
12. Scharstein, D., Taniai, T., and Sinha, S. N. (2017). Semi-global stereo matching with surface orientation priors. In: Proc. 2017 Int. Conf. 3D Vis. (3DV), 215–224.
13. Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., and Rother, C. (2008). A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. IEEE Trans. Pattern Anal. Mach. Intell. (PAMI), 30(6), 1068–1080.
14. Taniai, T., Matsushita, Y., Sato, Y., and Naemura, T. (2018). Continuous 3D label stereo matching using local expansion moves. IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI) 40(11), 2725–2739.
15. Wei, Y. and Quan, L. (2005). Asymmetrical occlusion handling using graph cut for multi-view stereo. In: Proc. IEEE Conf. Comp. Vis. Pattern Recog. (CVPR). 902–909.
16. Woodford, O., Torr, P., Reid, I., and Fitzgibbon, A. (2009). Global stereo reconstruction under second-order smoothness priors. IEEE Trans. Pattern Anal. Mach. Intell. (PAMI), 31 (12), 2115–2128.
17. Yoon, K. and Kweon, I. (2006). Adaptive support-weight approach for correspondence search. IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)28(4), 650–656.
18. Zabih, R. and Woodfill, J. (1994). Non-parametric local transforms for computing visual correspondence. In: Proc. Europ. Conf. Comp. Vis. (ECCV), 151–158.
19. Zbontar, J. and LeCun, Y. (2016). Stereo matching by training a convolutional neural network to compare image patches. J. Mach. Learn. Res. 17, 1–32.
20. Zhang, F., Prisacariu, V., Yang, R., and Torr, P.H. (2019). GA-Net: Guided Aggregation Net for End-To-End Stereo Matching. In: Proc. IEEE Conf. Comp. Vis. Pattern Recog. (CVPR).