

Evaluation of Stereo Algorithms for 3D Object Recognition

Federico Tombari Fabio Gori

Luigi Di Stefano

DEIS/ARCES - University of Bologna, Bologna, Italy

federico.tombari@unibo.it fabio.gori2@studio.unibo.it luigi.distefano@unibo.it

Abstract

This work aims at evaluating stereo matching algorithms in a 3D object recognition scenario, wherein objects have to be found and their 3D pose estimated efficiently and in presence of clutter and occlusions. Unlike all other surveys and evaluations of stereo methods, which compare accuracy based on a dataset of disparity maps with ground-truth, this work proposes an evaluation in terms of recognition ability which leverages on state-of-the-art approaches for 3D object recognition. The proposed evaluation methodology comprises a novel dataset characterized by realistic working conditions and compares state-of-the-art stereo algorithms potentially suitable to 3D object recognition applications.

1. Introduction

Stereo vision is a popular technique deployed to acquire geometrical (3D) information about the surrounding environment, which is often deployed in robotic applications, e.g. on board robotic platforms to help achieving the task of autonomous navigation and/or object manipulation. The main issue of stereo vision is still represented by stereo matching, a very challenging task counterpart by an active research community, with more than one hundred submitted proposals on the reference benchmark website [2] in the past decade, half of which were submitted just in the past 3 years. Given this vast literature and the huge amount of stereo matching algorithms available, the reader is often left with the question of which one would be the most suitable for his own application.

To help elucidating the above dilemma, several papers in recent literature proposed surveys and/or experimental evaluations focused on specific aspects of the vast stereo literature [21, 9, 11, 26, 25]. In particular, [21] proposed a taxonomy and an experimental evaluation of the whole class of stereo matching algorithms, classifying them into two main categories: *local* and *global*. Successively, [9] focused on local algorithms and presented an experimental

evaluation among methods suited to real-time implementation on a GPU. In addition, [11] presented a comparison between different matching functions aimed at robustness to photometric distortions and noise, while [26] proposed a classification and evaluation of cost aggregation strategies for local stereo algorithms. Recently, [25] proposed an evaluation of efficient and low-memory stereo algorithms (including an additional dataset wrt. those available on [2]) suited to robotic applications.

All the evaluations included in these works rely on comparing stereo algorithms in terms of accuracy of the retrieved disparity maps, i.e. given a stereo dataset and a labeled ground-truth disparity map, the number of wrong disparities is computed to rank the algorithms: this is the case of [21, 9, 11, 26, 25]. In addition, [9, 26, 25] also include an evaluation of methods in terms of their computational efficiency. Hence, all previous evaluations treat stereo as a general purpose 3D sensing technology, none of them addressing any specific application domain. Also, all previous evaluation works (except for [25]) were performed on the *Middlebury* datasets [2], which include images acquired under particularly good illumination conditions and by means of high-quality camera devices: this results in a particularly high image quality. This may easily not be the case in relevant application scenarios, e.g. mobile robotics, and it has been shown in [25] that the behavior of stereo algorithms tends to change significantly when practical nuisances and, in general, less ideal working conditions are accounted for.

In this work we propose a stereo matching algorithms evaluation focused on the specific field of 3D object recognition, a computer vision task aimed at detecting the presence and estimating the 3D pose of given objects, represented in the form of 3D models, from 3D data. More specifically, the main goal of this work is to determine which stereo algorithms deliver the best performance in the task of performing automatic 3D object recognition with clutter and occlusions under real working conditions. Moreover, another aspect that this work aims at investigating is whether the classical stereo evaluation based on accuracy of disparity maps and standard datasets allows for reliably

predicting the performance of the algorithms in the specific task of 3D object recognition in practical working conditions.

Therefore, unlike previous works, the considered stereo methods will not be ranked based on accuracy of disparity maps, but instead directly on 3D object recognition performance measured within a common state-of-the-art algorithmic framework. The evaluated methods have been selected among a pool of algorithms so as to span the main approaches for stereo matching, according to the established taxonomy proposed in [21]. All considered methods are state-of-the-art and computationally efficient, given the typical near real-time requirements of 3D object recognition applications, e.g. in the field of robotics. Therefore, methods requiring tens of seconds to several minutes to compute a single disparity map are not taken into consideration for the scope of this work. The pool of evaluated algorithms is presented in Section 2. Moreover, the evaluation is carried out on a novel, realistic dataset including typical nuisances of real application scenarios, such as uncontrolled ambient light, photometric distortions and artifacts, small defocus in the lenses, non-perfect rectification. The dataset, as well as the overall evaluation methodology, is described in Section 3.

Table 1. Summary of the evaluated stereo algorithms

| Method | Acronym |
|---------------------------|---------|
| Fixed Window | FW |
| ELAS [8] | ELAS |
| Fast Aggreg. [27] | FA |
| SemiGlobal [10] | SG |
| 2-pass Scanline Opt. [13] | 2SO |
| Dynamic Programming [21] | DP |
| Efficient BP [6] | EBP |
| Adaptive Cost DP [29] | ACDP |

2. Stereo Algorithms

As pointed out in [21], stereo matching algorithms can be divided into local and global. Methods belonging to the former category are generally more efficient than those of the latter.

Arguably, the most popular stereo matching algorithm is the *Fixed Window* (FW) or *Block-based* method. This approach assigns a disparity to a point of the reference image simply based on the similarity scores computed on a set of squared windows of fixed size centered at the candidates identified by the disparity range. By exploiting efficient computational schemes (i.e. Box Filtering [18], Summed Area Tables [4]) and/or by means of parallel processing, typically based on the SIMD instructions available nowadays in most general-purpose CPUs, implementations of this method can run in real-time, such as e.g. the stereo al-

gorithm included in the OpenCV library [1]. Nevertheless, due to its inherent simplicity, this baseline algorithm can not deal satisfactorily with common issues such as low-textured areas, depth discontinuities, occluded areas.

State-of-the-art local methods are generally based on a variable support approach [26], which allows for adaptively aggregating supporting pixels based on the characteristics of the local neighborhood instead of employing a fixed, squared window, as it the case of the FW algorithm. Several recent variable support-based methods can yield notably accurate disparity maps, but, unfortunately, tends to be quite slow (eg. several minutes for small sized images) [31]. Currently, the main research trend in variable support local methods deals with finding the best trade-off between accuracy and efficiency. In our evaluation we have considered a method based on the deployment of color segmentation to adapt the size of the support that attains a good accuracy-efficiency trade-off [27]. In addition, we have included a very recent efficient algorithm designed for large-scale stereo, presented in [8], that builds a prior over the disparity space by means of reliable correspondences (*support points*): this prior is then deployed to improve the accuracy of the stereo matching stage.

For what concerns the category of global methods, most approaches rely on formulations based on a global energy term that includes a smoothness constraint for better handling low-textured areas. Typical algorithms employed to determine approximate minima of the energy term are Graph Cuts [15] and Belief Propagation [14]. Although some research efforts [6, 33, 30] have recently made to improve the efficiency of global methods, they are still unable to achieve real-time performance, unless particular, hardware dependent accelerations are enforced (such as, e.g., GPU processing). In our approach we have included the method proposed in [6], which is a very popular global stereo method with publicly available source code wherein several optimizations, most notably a pyramidal approach, are exploited to significantly improve the efficiency of the algorithm.

Stereo matching approaches based on Scanline Optimization (SO) and Dynamic Programming (DP) trade accuracy for efficiency by minimizing a global energy term over image scanlines. More advanced schemes based on DP and SO have been proposed [17, 10], so as to deploy several scanlines simultaneously and build up a cost function which is more “global” (trying to fetch information from different areas of the image). In our evaluation we have included the standard DP algorithm [21] as well the SemiGlobal algorithm proposed in [10]. In addition, we have considered a two-pass SO algorithm similar to that proposed in [13], which is based on the minimization of an aggregated cost computed by applying SO on two scanlines, one horizontal from left to right, and the other one horizontal again but

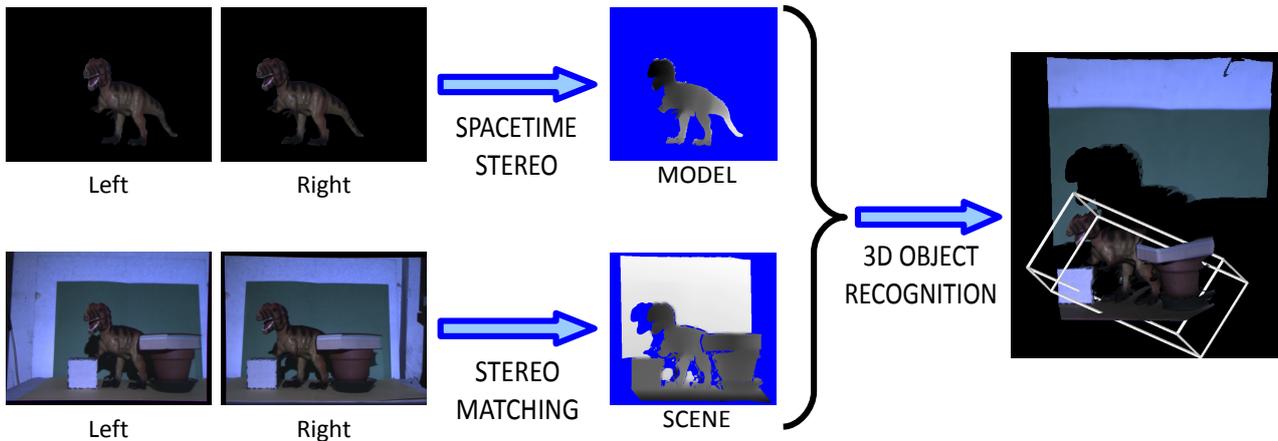


Figure 1. Flow diagram of the evaluation methodology.

from right to left.

An additional set of methods is represented by those deploying an adaptive support to compute a local aggregation cost in conjunction with a global approach embedding a smoothness constraint to improve the local cost [32, 13, 29]. We have included in our evaluation the efficient method proposed in [29] which relies, in the local aggregation step, on a simplification of method proposed in [31] consisting in computing the cost over vertical columns only [29], and then on deploying DP for the purpose of energy minimization.

A summary of the methods included in our comparison, together with the respective acronym that will be used throughout the paper, is presented in Table 1.

3. Evaluation framework

3.1. 3D Pipeline

The proposed evaluation relies on a typical 3D pipeline for object recognition based on 3D feature matching, which is sketched in Fig. 1. Feature matching is a common approach for object recognition that has been particularly successful in the past years and has been used for both 3D point clouds [12, 28, 19, 20, 35] as well as images [16, 3]. First of all, a 3D mesh of each model and the scene is attained by means of a 3D processing technique. In our case, the considered stereo matching algorithm computes a disparity map of the current scene from a pair of rectified stereo images. As for the models, instead, their disparity map is attained by the Spacetime Stereo (STS) technique [5, 34] followed by the 3D processing pipeline described in [24], which provides for highly accurate disparity maps. This is done to re-create a scenario wherein the characteristics of the 3D data related to the models are different from those of the scene due to different acquisition sensors being used. In

such a scenario, the models can typically be reconstructed much more accurately than the scene, as models can be acquired off-line.

Then, from each disparity map a 3D mesh can be quickly obtained [24]. Successively, a detector extracts a set of 3D *features*, which represents 3D points generally characterized by distinctiveness and repeatability. Each of these features is then *described* by means of a 3D descriptor [12, 28], i.e. a vector is computed that encodes some geometrical characteristics of the neighborhood of the 3D point. As this is done for both the model and for the scene, model features can then be matched to scene features based on descriptors, with the best matching feature pairs yielding feature correspondences. Finally, these correspondences are passed to a *geometric validation* algorithm [12, 23, 20], which aims at filtering out outliers by enforcing a consensus generally based on geometrical constraints.

As far as the specific techniques employed in this work are concerned, first 3D features are randomly sampled from the 3D mesh as done in [12]. Then, a state-of-art descriptor, namely SHOT [28], is deployed to yield a description of each 3D feature. Finally, the Hough-based object geometric validation algorithm proposed in [23] is used to determine the presence of a model in a scene.

3.2. Dataset

The proposed evaluation requires a dataset with specific characteristics: we need stereo-pairs of models and scenes acquired under realistic conditions as well as the ground-truth concerning the 3D position of each model in the scene. At the best of our knowledge, there is no publicly available dataset with the required characteristics, thus we acquired such a dataset in our lab specifically for this evaluation. The dataset includes 4 models (referred to as *Dino*, *Mario*, *Rabbit*, *Bust* and shown in Figure 2, top) and 55 scenes, divided



Figure 2. Top: the 4 models used to create the scenes. Bottom: 4 selected scenes belonging to the dataset.

into two subsets: a *Test* dataset, made out of 27 scenes, and a *Tuning* dataset, comprising 28 scenes and used for parameters tuning (see next). Overall, the *Test* dataset includes a total of 63 object recognition experiments (27 positives, 36 negatives). Figure 2, bottom shows the reference (i.e. left) stereo image concerning 4 scenes belonging to the dataset and characterized by different occlusion levels.

As mentioned previously, model data are different (i.e. more accurate) than scene data, for each model has been acquired through a stereo sequence of 100 frames with a variable random pattern projected on each frame, so as to reconstruct it by means of the Spacetime Stereo (STS) technique [5, 34] and the 3D acquisition pipeline described in [24]. Differently, each scene is represented only by a stereo pair, and will be processed by each of the evaluated stereo matching algorithms. In each scene, one or more models appear at different levels of occlusion and within clutter. Occlusions are created by means of other objects and cover several levels of moderate occlusion (ranging from 5% to 50%).

To give an idea of the challenge represented by the proposed dataset, as well as to compare qualitatively the output of the stereo algorithms on a typical scene of the dataset, Figure 3 shows the range maps yielded by each of the evaluated algorithm on one scene of the dataset. Also, on top, the Figure shows the reference intensity image and the range map obtained using the STS technique. It is quite clear that global algorithms have better performance in terms of accuracy over the local ones. Also, the STS map is notably more accurate than those attainable by stereo algorithms.

3.3. Implementation

As for EBP and ELAS, we have used the publicly available code provided by the authors of the original papers. We have instead deployed our own implementations of methods

FW, FA, SG, SO, DP, ACDP. To the best of our knowledge, no evaluated algorithms employs any kind of optimization such as e.g. SIMD instructions: this allows for a fair comparison also in terms of computational efficiency. Given the real-world acquisition conditions for the dataset, a pre-processing stage is applied to images. This stage relies on the subtraction to each pixel of both the left and right images of the mean value computed over a small, squared window centered at the pixel.

Since also a post-processing stage is commonly deployed when working with stereo images, we will evaluate the performance of the algorithms with a simple cascade of post-processing filters composed of three stages. The first is the Left-Right Consistency check [7], a simple filter that eliminates non-consistent disparities between the two stereo views and requires the computation of both the left and the right disparity maps. The other two post-processing stages are, in order, a median filter, to remove disparity outliers, and an adaptive mean (i.e. *bilateral*) filter [22] applied to the disparity map, so as to smooth out the object surface while preserving depth borders.

3.4. Evaluation metrics

Each object recognition experiment can be seen as a binary classification test, where the presence of a model in a scene has to be classified either as positive or negative. When the model is present in the current scene, a true positive is scored if the algorithm detects its presence and the alignment error between the model and its instance in the current scene (computed according to the available ground-truth) is below a pre-determined threshold (in our experiments it was set to 5 times the average mesh resolution of the dataset). A false positive is scored if the alignment error is higher than the threshold or if the model is detected in a scene where it is actually absent. Hence, it is possible

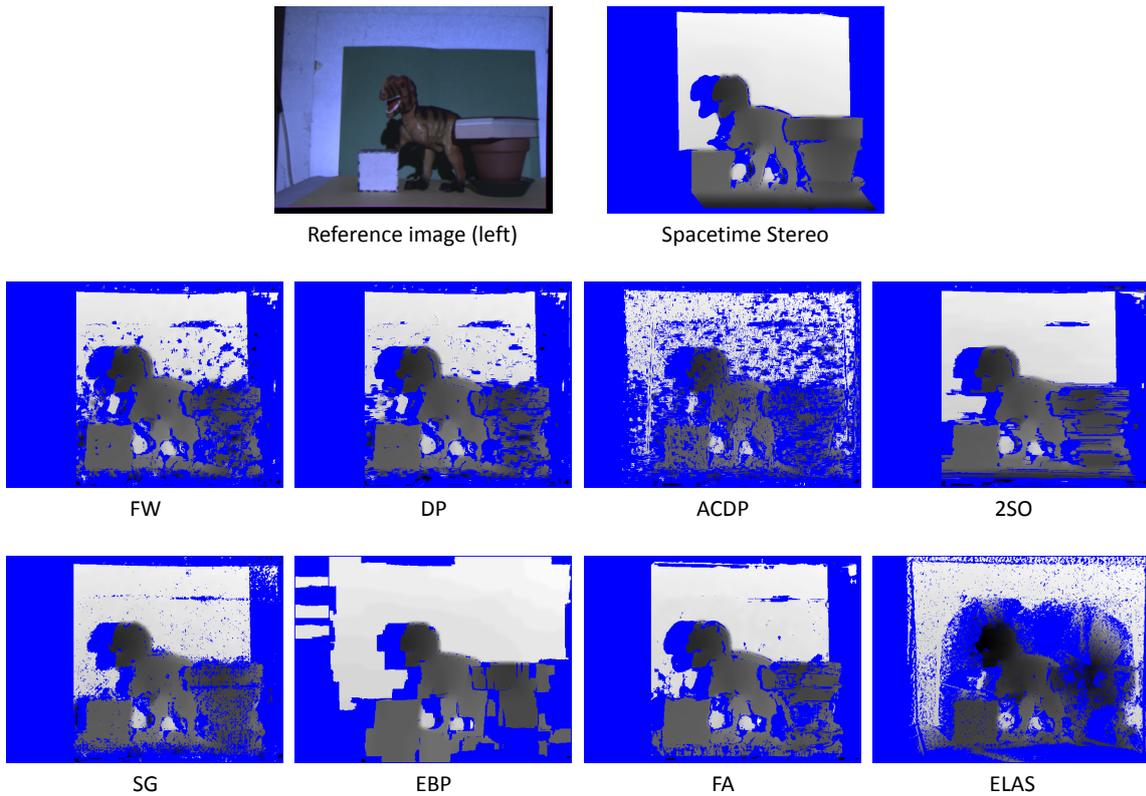


Figure 3. Range maps yielded by the evaluated algorithms (see Table 1) compared to that achievable by STS (Top right). Top left: the reference intensity image.

to evaluate the performance of the algorithms in terms of *Precision-Recall curves*, spanning over the threshold used to detect the presence of an object over the final set of geometrically validated correspondences.

3.5. Parameter selection

Since the proposed pipeline includes several stages, each characterized by a certain number of parameters, we have performed an accurate parameter selection, as outlined in the following.

As for stereo algorithms, we have experimented two different tuning strategies, which will lead to two different experiments in the next Section (*Exp. A* and *Exp. B*). The first strategy (*Exp. A*) relies on choosing the parameter values that directly maximize the recognition capabilities over the *Tuning* dataset, i.e. in terms of *Precision-Recall curves*. Additionally, we were interested in comparing the performance of the algorithms with those yielded by means of a different, more “standard” tuning approach, that is, choosing the parameter values in order to maximize the accuracy provided by the stereo algorithms over a dataset of disparity maps (*Exp. B*). It is worth pointing out that the latter tun-

ing approach is usually deployed to select the parameters of stereo algorithms by the research community (e.g. see the Middlebury website [2]), but also within some application scenarios. Hence, we selected this different set of parameters by tuning them on the realistic stereo dataset proposed in [25] (more details about the error measure and the dataset can be found in [25]). As it will be shown in the experimental section, the dependence of the behavior of the algorithms on the tuning strategies is relevant.

Since the remaining stages of the pipeline (feature detection, description, matching and validation) other than stereo matching are in common to all stereo algorithms, we performed a single parameter tuning stage for these stages separated from that provided for stereo algorithms, so not to be influenced by the characteristics of some particular stereo algorithm. Specifically, we tuned the parameters of the SHOT feature descriptor and of the Hough-based geometric validation algorithm over a tuning dataset composed of the same 4 models used in the testing set but of 26 different tuning scenes, by selecting the parameters yielding the best performance in terms of *Precision-Recall curves*. To render this tuning stage completely independent from the

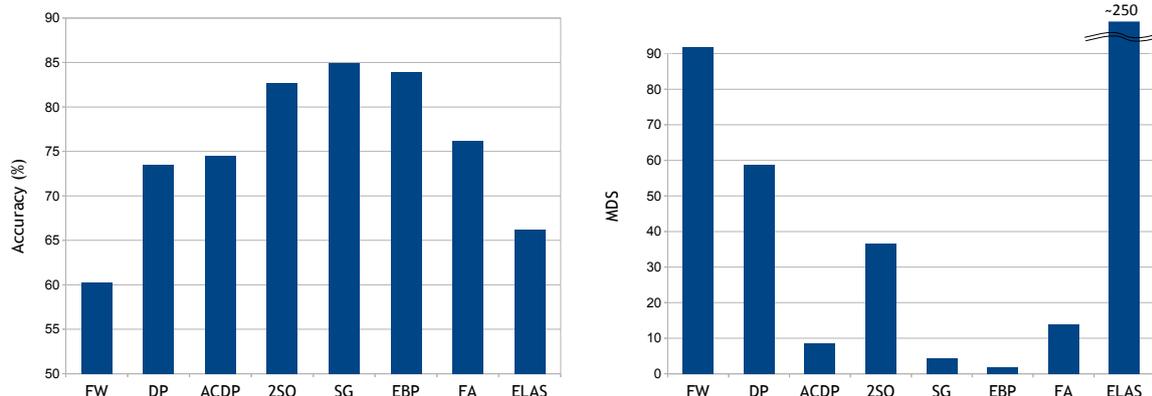


Figure 4. Results in terms of accuracy (left) and efficiency (MDS, right) of the evaluated algorithms on the stereo dataset proposed in [25].

stereo matching algorithm used in our evaluation, each disparity map of the models and scenes used for the evaluation was obtained by means of the STS technique [5, 34] and the 3D acquisition pipeline described in [24].

4. Experimental results

This section discusses the experimental results yielded by our evaluation. Results are divided into two different object recognition experiments, each relative to one of the two tuning procedures described in Subsection 3.5. In particular, the first Experiment (*Exp. A*) refers to the tuning based on maximizing recognition capabilities, while the second one (*Exp. B*) refers to the tuning based on maximization of accuracy on disparity maps. A first outcome of the latter tuning procedure is shown in Figure 4, which reports the best (over parameter values variations) disparity map accuracy achieved by each algorithm. In addition, the Figure shows the efficiency of each algorithm measured in millions of disparities per second (MDS). As expected, global algorithms yield highest accuracy. In particular SG turns out to be the best algorithm with a 85% of accuracy, closely followed by EBP and 2SO, while FW reaches 60% accuracy only. ELAS is found to be really efficient, although it does not reach the same accuracy as global algorithms. Overall, these results are consistent with what should be found in a classic evaluation based on disparity maps quality.

Figure 5 shows the results of both Experiment A and Experiment B conducted on the *Test* dataset. Here all the employed stereo matching algorithms are compared in terms of *Precision-Recall curves*. As for Exp. A, it can be seen that, overall, the algorithms do not reach particularly high performance, as for any Precision value no algorithm reaches more than 0.52 in Recall. This occurred despite the fact that occlusions were limited to a maximum of 50% per model in each scene, proving that 3D object recognition is a very

difficult task when near real-time stereo matching is used to acquire the 3D scene. In both experiments the best performance is achieved by SG, confirming the good accuracy already achieved in the retrieved disparity maps (see Fig.4). However, other algorithms with similar scores in terms of disparity map accuracy, namely EBP and 2SO, yielded a poor performance in terms of *Precision-Recall curves* despite the good results in Fig.4. After a careful analysis, we observed that this effect is due to the regularization criteria included in these algorithms, which tends to award smooth background surfaces, with this sometimes distorting the reconstructed shapes of foreground objects in a way that deteriorates the distinctiveness of the local features upon which the adopted 3D object recognition pipeline relies. For the same reason, DP and ACDP also do not yield good performance, their recognition capabilities being lower than that of the FW algorithm. Finally, local algorithms such as ELAS and FA provide for interestingly good performance, ranking just after SG, with ELAS being the overall best performer at high Precision values while FA performing notably better at lower Precision values.

Furthermore, comparing the two experiments, we can see that Exp. A curves indicate a significant improvement over Exp. B ones, with all algorithms performing better in Exp. A and this being more evident for FW, DP, ACDP and FA. As for DP and ACDP, the disparity-based tuning process tend to enforce an high degree of smoothness on the reconstructed surface that render local features poorly distinctive, at the point that is turns out impossible to perform any recognition. This aspect is better highlighted in Figure 6, which presents the comparison between the results coming from the two tuning strategies for each stereo matching algorithm. These results suggest that the classic disparity-based tuning procedure is not suitable for a 3D object recognition framework relying on local features. In fact, given the structure of currently available stereo dataset, such a tuning

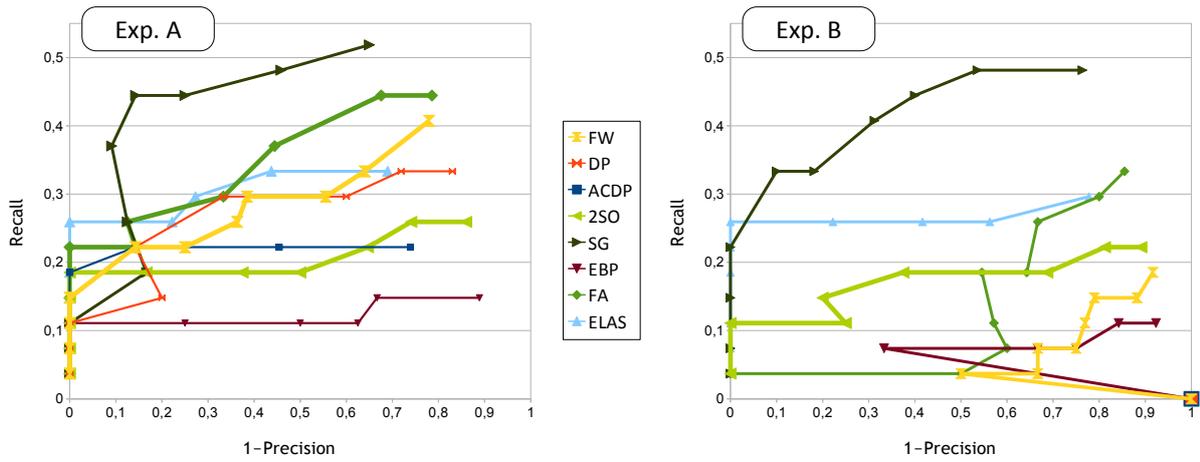


Figure 5. Precision-Recall reported by the evaluated algorithms on the two experiments (Exp. A and B).

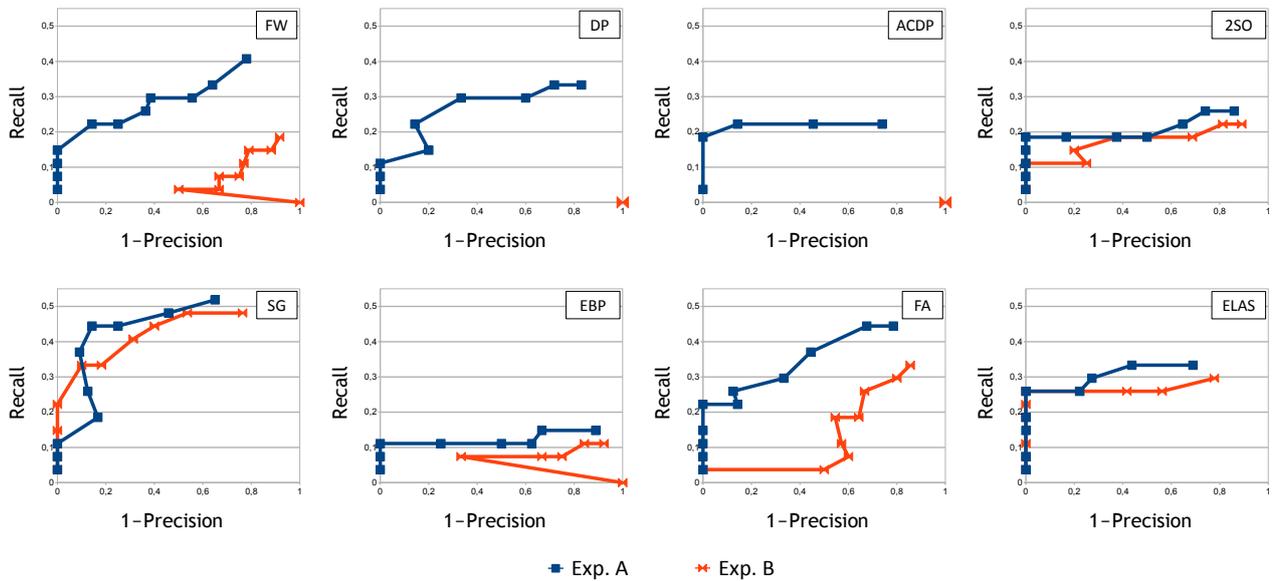


Figure 6. Difference in performance of the evaluated algorithms using as tuning criterion the maximization of the recognition capabilities (Exp. A) and the minimization of the disparity error (Exp. B).

approach because tends to privilege the smoothness of reconstructed surfaces, which is good to improve the global accuracy of disparity maps but tends to wash out the differences among local shape features, often making them not enough distinctive to allow for successful object recognition.

5. Conclusion and future work

We have proposed a comparison between near real-time stereo matching algorithms specifically focused on 3D ob-

ject recognition. To this purpose, we have assessed the performance of several state-of-the-art stereo methods representative of different approaches within a realistic scenario. Thanks to the comparison between two different parameters tuning strategies, i.e. the standard one and an application-specific one, it turned out that only a subset of the best performing algorithms in terms of disparity map accuracy are actually suitable to perform stereo-based object recognition, which still remains a very challenging task. Since this work has taken into account only approaches based on local features and regional descriptors, future work is represented

by the extension of the comparison to the case of 3D object recognition based on global methods. Also, the ideas at the base of our work can be deployed to evaluate stereo algorithms with a focus on other relevant applications such as point clouds registration for 3D reconstruction and SLAM.

References

- [1] <http://opencv.willowgarage.com>. 2
- [2] Middlebury stereo evaluation. <http://vision.middlebury.edu/stereo>. 1, 5
- [3] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. Surf: Speeded up robust features. *CVIU*, 110(3):346–359, 2008. 3
- [4] F. Crow. Summed-area tables for texture mapping. *Computer Graphics*, 18(3):207–212, 1984. 2
- [5] J. Davis, D. Nehab, R. Ramamoorthi, and S. Rusinkiewicz. Spacetime stereo : A unifying framework for depth from triangulation. *PAMI*, 27(2):1615–1630, 2005. 3, 4, 6
- [6] P. Felzenszwalb and D. Huttenlocher. Efficient belief propagation for early vision. In *Proc. CVPR 2004*, 2004. 2
- [7] P. Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. *Machine Vision and Applications*, 6(1):35–49, 1993. 4
- [8] A. Geiger, M. Roser, and R. Urtasun. Efficient large-scale stereo matching. In *Proc. Asian Conference on Computer Vision*, November 2010. 2
- [9] M. Gong, R. Yang, W. Liang, and M. Gong. A performance study on different cost aggregation approaches used in real-time stereo matching. *Int. Journal Computer Vision*, 75(2):283–296, 2007. 1
- [10] H. Hirschmuller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *Proc. Conf. on Computer Vision and Pattern recognition (CVPR 2005)*, volume 2, pages 807–814, 2005. 2
- [11] H. Hirschmuller and D. Scharstein. Evaluation of cost functions for stereo matching. In *Proc. Conf. on Computer Vision and Pattern Recognition (CVPR 2007)*, volume 1, pages 1–8, 2007. 1
- [12] A. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3d scenes. *PAMI*, 21(5):433–449, 1999. 3
- [13] J. Kim, K. Lee, B. Choi, and S. Lee. A dense stereo matching using two-pass dynamic programming with generalized ground control points. In *Proc. Conf. on Computer Vision and Pattern Recognition (CVPR 2005)*, pages 1075–1082, 2005. 2, 3
- [14] J. Kim and J. Pearl. A computational model for combined causal and diagnostic reasoning in inference systems. In *Proc. 8th Int. J. Conf. on Artificial Intelligence (IJCAI)*, pages 190–193, 1983. 2
- [15] V. Kolmogorov and R. Zabih. What energy functions can be minimized using graph cuts? *Trans. PAMI*, 2002. 2
- [16] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60:91–110, 2004. 3
- [17] G. M. Bleyer, M. Simple but effective tree structures for dynamic programming-based stereo matching. In *Proc. Int. Conf. on Computer Vision Theory and Applications (VIS-APP)*, volume 2, 2008. 2
- [18] M. Mc Donnel. Box-filtering techniques. *Computer Graphics and Image Processing*, 17:65–70, 1981. 2
- [19] A. Mian, M. Bennamoun, and R. Owens. A novel representation and feature matching algorithm for automatic pairwise registration of range images. *IJCV*, 66(1):19–40, 2006. 3
- [20] A. S. Mian, M. Bennamoun, and R. A. Owens. On the repeatability and quality of keypoints for local feature-based 3d object retrieval from cluttered scenes. *IJCV*, 89(2-3):348–361, 2010. 3
- [21] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. Jour. Computer Vision*, 47(1/2/3):7–42, 2002. 1, 2
- [22] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proc. CVPR*, 1998. 4
- [23] F. Tombari and L. Di Stefano. Object recognition in 3d scenes with occlusions and clutter by hough voting. In *Proc. 4th Pacific-Rim Symp. on Image and Video Technology (PSIVT 10)*, 2010. 3
- [24] F. Tombari, L. Di Stefano, S. Mattoccia, and A. Mainetti. A 3d reconstruction system based on improved spacetime stereo. In *Proc. Int. Conf. on Control, Automation, Robotics and Vision (ICARCV 2010)*, 2010. 3, 4, 6
- [25] F. Tombari, S. Mattoccia, and L. Di Stefano. Stereo for robots: quantitative evaluation of efficient and low-memory dense stereo algorithms. In *Proc. Int. Conf. on Control, Automation, Robotics and Vision (ICARCV 2010)*, 2010. 1, 5, 6
- [26] F. Tombari, S. Mattoccia, L. Di Stefano, and E. Addimanda. Classification and evaluation of cost aggregation methods for stereo correspondence. In *Proc. CVPR*, 2008. 1, 2
- [27] F. Tombari, S. Mattoccia, L. Di Stefano, and E. Addimanda. Near real-time stereo based on effective cost aggregation. In *Proc. Int. Conf. on Pattern Recognition (ICPR 08)*, 2008. 2
- [28] F. Tombari, S. Salti, and L. Di Stefano. Unique signatures of histograms for local surface description. In *ECCV*, 2010. 3
- [29] L. Wang, M. Liao, M. Gong, R. Yang, and D. Nister. High-quality real-time stereo using adaptive cost aggregation and dynamic programming. In *Proc. 3rd Int. Symposium 3D Data Processing, Visualization and Transmission (3DPVT'06)*, pages 798–805, 2006. 2, 3
- [30] Q. Yang, L. Wang, and N. Ahuja. A constant-space belief propagation algorithm for stereo matching. In *Proc. CVPR 2010*, 2010. 2
- [31] K. Yoon and I. Kweon. Adaptive support-weight approach for correspondence search. *IEEE Trans. PAMI*, 28(4):650–656, 2006. 2, 3
- [32] K. Yoon and I. Kweon. Stereo matching with symmetric cost functions. In *Proc. Conf. on Computer Vision and Pattern Recognition (CVPR 2006)*, volume 2, pages 2371 – 2377, 2006. 3
- [33] T. Yu, R. Lin, B. Super, and B. Tang. Efficient message representations for belief propagation. In *Proc. ICCV 07*, 2007. 2
- [34] L. Zhang, B. Curless, and S. Seitz. Spacetime stereo: Shape recovery for dynamic scenes. In *CVPR*, 2003. 3, 4, 6
- [35] Y. Zhong. Intrinsic shape signatures: A shape descriptor for 3d object recognition. In *ICCV-WS:3DRR*, 2009. 3