An Improved Evaluation Method for Optical Flow of Angle Error

Zhen CHEN[1,2,*],Ren-zhi QIN[1],Cong-xuan ZHANG[1,2],Ding-ding KONG[2] and

Xuan SHANG[2]

[1]Key Laboratory of Nondestructive Testing, Ministry of Education, Nanchang, China 330063

[2]School of Measuring and Optical Engineering, Nanchang Hangkong University, Nanchang, China, 330063

[*]Corresponding author

Email: dr_chenzhen@163.com

Keywords: Optical flow, Corrected, Middlebury benchmark, Angle error, Middlebury database.

Abstract. To address the challenge of evaluating modern optical flow computation technology, this paper proposes a corrected evaluation methodology for optical flow of angle error based on the popular Middlebury benchmark. Firstly, a corrected evaluation benchmark is proposed based on the Middlebury benchmark. for flow error statistics, a mathematical expectation is employed instead of the current mathematical average as the expectation can reflect a more appropriate error distribution for the flow result. Secondly, several typical methods of optical flow and four classical test sequences chosen from the Middlebury database are employed to show the improvement attained by using the proposed evaluation methodology. Thirdly, evaluation results of the test methods with both the proposed and Middlebury benchmarks are provided in the experimental section. Finally, the comparison result indicates that the proposed evaluation benchmark can reflect the flow result performance more objectively and appropriately.

Introduction

In the last decade, the accuracy, robustness, and efficiency of optical flow estimation have undergone significant development owing to improvements in the level of computer software and hardware. In 1994,Barron[1] proposed an evaluation system which contained mainly the average angle error, the standard deviation, and the density of the computed optical flow result. To overcome the lack of ground truth data sets for complex scenarios, McCane[2] presented a method for generating motion fields from real sequences containing polyhedral objects and provided the Otago[3] optical flow evaluation database.

With the development of optical flow algorithms, test image sequences in the Otago database could not satisfy evaluation requirements for the next generation of optical flow algorithms: a modern optical flow evaluation database should be closer to real scenarios that involve non-rigid motion, larger motion ranges, more realistic textures, independent motion, and more complex occlusions. To promote rapid progress in optical flow estimation, researchers at Middlebury College produced a new set of benchmarks for modern optical flow algorithms and made the test image sequence data freely available on the web at http://vision.middlebury.edu/flow/[4]in 2007, providing four types of image sequence data to test different aspects of optical flow algorithms: (1) sequences containing non-rigid motion where the ground truth flow is determined by tracking a hidden fluorescent texture, (2) realistic synthetic sequences with the ground truth, (3) modified stereo

sequences with the ground truth, and (4) high-frame-rate sequences without the ground truth. Figure 1 shows several representative image sequences with the ground truth from the Middlebury database.

Figure 1. From top to bottom: the reference frame of sequence and the corresponding ground truth.From left to right: Dimetrodon ,Grove3, Rubberwhale. and Urban2.

Current Evaluation Methodology

Angular Error

In Barron's research[5], the angular error was proposed as the spacing angle between the estimated optical flow and the ground truth,which can be expressed as followed:

$$AE = \arccos\left(\frac{u \times u_{GT} + v \times v_{GT} + \delta}{\sqrt{u^2 + v^2 + \delta}\sqrt{u_{GT} + v_{GT} + \delta}} \right) \tag{1}$$

where $(u,v)^T$ denotes the estimated optical flow vector and $(u_{GT}, v_{GT})^T$ denotes the ground truth of the optical flow. It is not difficult to see that the angular error formula added a third coordinate δ (δ was equal to $\delta = 1$ after the work of McCane[2]) to the estimated optical flow and the ground truth, which significantly lowered angle differences for large optical flows.

Statistics

In Barron's research[5], the averages (Avg) and standard deviations (STD) were employed as statistical measures of optical flow errors through which angle error statistics can be computed as follows:

$$Avg = \frac{1}{N}\sum_{i=1} AE(i), STD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(AE(i) - AAE)^2} \tag{2}$$

To demonstrate the performance of a test algorithm comprehensively, Baker[6] proposed a novel statistical measure method which contained mainly robustness and accuracy statistics. Robustness statistics (RX) denotes the percentage of pixels that have an error measure above X while accuracy statistics (AX) denotes the accuracy of the error measure at the Xth percentile after sorting the errors from lowest to highest. Table 1 shows the statistical measure used in the Middlebury dataset proposed by Baker[6].

Table 1. The statistical measure used in the Middlebury dataset

AE	Avg	STD	R2.5	R5.0	R10.0	A50	A75	A95

Currently, the statistical measure of the Middlebury benchmark is the most widely acknowledged evaluation method for optical flows because it gives a comprehensive appraisal of optical flow algorithm performance.

Corrected Evaluation Methodology

Corrected Angular Error

In this paper, a corrected error measure for the angle that only reflects the angular deviation of the optical flow is proposed as follows:

$$AE = \begin{cases} \arccos(v \cdot v_{GT}), & |v| \neq 0 \, \& \, |v_{GT}| \neq 0 \\ 180^0, & |v| = 0 \, | \, |v_{GT}| = 0 \end{cases} \tag{3}$$

Corrected Statistics

According to the unequal probability distribution of the error of an optical flow, the average and the corresponding standard deviation are not suitable as statistics measures because they cannot reflect the error distribution accurately. To furnish more appropriate statistics measures, the expectation (EX) is employed instead of the average in the proposed method as below:

$$EX = \sum_{i=1}^{N} x_i p_i \tag{4}$$

where N denotes the count of the error, x_i denotes the value of the error, and denotes the corresponding probability of the error, which can be computed from the error of the optical flow. with the proposed statistics measure of the expectation, the corresponding standard deviation can be rewritten as follows:

$$EXSTD = \sqrt{\sum_{i=1}^{N} (x_i - EX)^2 * p_i} \tag{5}$$

To provide a thorough evaluation, the robust statistics index RX and the accuracy measure index AX derived from the Middlebury benchmark were employed to evaluate the robustness of the optical flow in the proposed method. RX denotes the percentage of pixels that have an error measure above X and AX denotes the accuracy of the error measure at the Xth percentile. In contrast to the Middlebury benchmark, according to the robust statistics index RX, for the angle error (AE) R2.5, R5.0, and R10.0 were computed. According to the accuracy measure index AX, for the flow errors,A95 is computed to show the robustness of the accuracy. With the corrected error measures and statistics proposed above, the improved statistics measures can be rewritten as shown in Table 2.

Table 2. The improved statistic measures

AE	EX	EXSTD	R2.5	R5.0	R10.0	A95

Experiments and Analysis

In this section, the proposed evaluation methodology for optical flow is discussed. The experiments focused mainly on differences between the proposed benchmark and the Middlebury benchmark. However, the comparison of optical flow results of the adopted methods is not the highlight. Owing to the large number of measures for the proposed evaluation methodology, several selective methods of optical flow were employed to show the evaluated results of the proposed benchmark and the Middlebury benchmark, and specific analysis was attached to each evaluation.

Evaluation Methods

For intuitive comparison, the selective methods included mainly LDOF[7], SODOF[8], Correlation Flow[9], and Classic+nl[10].In addition, the computed parameters of the evaluation methods were set without careful consideration, so the parameters were set mainly as fixed values proposed by their respective articles, which may not be optimal for each test. Although some discussion of the performance of the above methods is provided, the ranking of the evaluation methods not of interest as the highlight of this paper is to emphasize the differences between the proposed evaluation benchmark and the current popular Middlebury benchmark.

Angle Error

For the analysis of angle errors for each sequence, the statistical results of angle errors of the All masks show the performance of the evaluation methods with various computing challenges.

With the statistics results of the angle error in Table 3 and Table 4, the best result of each statistics item for one sequence is highlighted with red and bold text. the first thing to note is that the distribution of the best results for the proposed benchmark is mainly unchanged compared with the Middlebury benchmark except for minor changes in a few statistics items, indicating that the proposed evaluation methodology would not influence the existing rank of the Middlebury benchmark drastically,the statistic results of the AE with the Middlebury benchmark are as follows.

Table 3. The statistic results of the AE with the Middlebury benchmark

METHOD			Avg	STD	R2.5	R5	R10	A50	A75	A90
LDOF	Dimetrodon	ALL	2.182	2.441	0.701	0.919	0.980	1.451	2.816	6.607
	Grove3		4.476	12.655	0.733	0.840	0.915	1.075	2.728	16.542
	RubberWhale		3.975	11.748	0.794	0.878	0.928	1.039	2.051	15.736
	Urban2		2.377	8.081	0.817	0.920	0.959	0.495	1.889	8.002
SODOF	Dimetrodon	ALL	3.838	3.767	0.469	0.740	0.943	2.711	5.122	10.668
	Grove3		6.548	13.752	0.511	0.700	0.859	2.400	6.239	24.941
	RubberWhale		4.278	9.512	0.577	0.849	0.932	2.143	3.576	13.209
	Urban2		2.944	7.299	0.715	0.860	0.944	1.025	2.940	10.961
Correlation Flow	Dimetrodon	ALL	5.004	6.234	0.474	0.694	0.873	2.688	6.156	18.497
	Grove3		4.342	10.759	0.622	0.817	0.929	1.726	3.841	12.986
	RubberWhale		2.474	5.639	0.741	0.932	0.979	1.531	2.551	6.059
	Urban2		3.419	8.010	0.691	0.824	0.925	1.048	3.343	13.646
Classic+nl	Dimetrodon	ALL	2.306	2.309	0.681	0.889	0.989	1.490	3.115	6.914
	Grove3		4.207	13.645	0.770	0.878	0.932	1.026	2.308	13.464
	RubberWhale		2.516	8.643	0.867	0.940	0.967	1.002	1.668	6.056
	Urban2		1.798	6.717	0.847	0.939	0.977	0.513	1.491	5.804

More importantly, the statistics results of EX indicate that the expectation cannot be replaced by the average because the expectation reflects the distribution of the error; the statistical results of Dimetrodon, Grove3, and Urban2 in Table 4 demonstrate that pixels with small angle error are of large probability in these sequences for the evaluation methods and pixels with large angle error are of large probability in the Rubberwhale sequence, the statistic results of the AE with the proposed benchmark are as follows.

Table 4. The statistic results of the AE with the proposed benchmark

METHOD			EX	EXSTD	R2.5	R5	R10	A95
LDOF	Dimetrodon	ALL	1.748	2.649	0.760	0.913	0.974	7.109
	Grove3		4.208	13.345	0.739	0.847	0.920	16.246
	RubberWhale		5.486	21.529	0.805	0.885	0.924	20.152
	Urban2		2.704	10.778	0.807	0.897	0.955	9.239
SODOF	Dimetrodon	ALL	3.326	4.327	0.520	0.758	0.934	11.545
	Grove3		6.508	15.617	0.555	0.715	0.857	28.159
	RubberWhale		5.159	17.792	0.625	0.834	0.927	15.374
	Urban2		3.152	10.584	0.755	0.860	0.937	12.392
Correlation Flow	Dimetrodon	ALL	4.741	7.282	0.515	0.707	0.864	21.064
	Grove3		4.218	12.050	0.642	0.812	0.922	14.436
	RubberWhale		2.519	9.631	0.735	0.915	0.969	7.114
	Urban2		4.219	13.843	0.720	0.827	0.909	16.659
Classical+nl	Dimetrodon	ALL	1.952	2.433	0.696	0.880	0.980	7.394
	Grove3		4.067	14.419	0.775	0.872	0.931	13.963
	RubberWhale		3.128	15.310	0.860	0.936	0.962	6.746
	Urban2		2.139	9.360	0.851	0.919	0.965	7.743

For standard deviation, the numeric values of EXSTD for the proposed benchmark are generally larger than the numeric values of STD in the Middlebury benchmark, and the small numeric values of EXSTD indicate the error fluctuation of the evaluation methods as EXSTD expresses the real distribution of the angle errors. For the robust statistics index RX and the accuracy measure index AX, the statistical results show the robustness and convergence of the evaluation methods, and the accuracy measure index AX was decreased to one index A95 compared with the Middlebury benchmark.

Conclusions

In this paper, an improved evaluation methodology for optical flow of angle error based on the current evaluation benchmark was presented. For the problem of objective evaluation, this paper proposed an improved evaluation benchmark by modifying the metrics and statistics of optical flow error. In addition.For flow error statistics, we offered a mathematical expectation and corresponding standard deviation instead of the currently popular mathematical average and standard deviation, and the proposed expectation provides a more appropriate error distribution for the flow result. In the experimental section, we conducted several typical methods of optical flow with synthetic and real sequences chosen from the Middlebury database, and we provided the evaluation results of the test methods with both the proposed benchmark and the Middlebury benchmark. The comparison results indicated that the proposed evaluation benchmark reflected the performance of the flow result more objectively and appropriately, even though the ranking of the evaluation methods by the proposed benchmark was approximately the same as for the Middlebury benchmark.

References

[1] Barron J L, Fleet D J, Beauchemin S S. Performance of optical flow techniques[J]. International journal of computer vision, 1994, 12(1): 43-77.

[2] Bruhn A, Weickert J, Schnörr C. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods[J]. International Journal of Computer Vision, 2005, 61(3): 211-231.

[3] http://of-eval.sourceforge.net/.

[4] http://vision.middlebury.edu/flow/.

[5] Barron J L, Fleet D J, Beauchemin S S. Performance of optical flow techniques[J]. International journal of computer vision, 1994, 12(1): 43-77.

[6] Baker S, Scharstein D, Lewis J P, et al. A database and evaluation methodology for optical flow[J]. International Journal of Computer Vision, 2011, 92(1): 1-31.

[7] Brox T, Malik J. Large displacement optical flow: descriptor matching in variational motion estimation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(3): 500-513.

[8] Ayvaci A, Raptis M, Soatto S. Sparse occlusion detection with optical flow[J]. International Journal of Computer Vision, 2012, 97(3): 322-338.

[9] Drulea M, Nedevschi S. Motion estimation using the correlation transform[J]. IEEE Transactions on Image Processing, 2013, 22(8): 3260-3270.

[10]Sun D, Roth S, Black M J. Secrets of optical flow estimation and their principles[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2010: 2432-2439.