Advances in Computer Science Research (ACRS), volume 54
International Conference on Computer Networks and Communication Technology (CNCT2016)

An Improved Evaluation Method for Optical Flow of Endpoint Error

Cong-xuan ZHANG[1,2,*], Ling-ling ZHU[1,2],Zhen CHEN[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

E-Mail: zcxdsg@163.com

Keywords: Optical flow, Evaluation, Benchmark, Corrected.

Abstract. A corrected evaluation benchmark is proposed based on the Middlebury benchmark. The major contributions of the presented benchmark include the following: error metrics of endpoint error are modified, and the corrected metrics can decrease the detrimental influence of human factors and reflect the actual flow evaluation performance. In addition, the metric of normalization endpoint error is projected to reveal the relative error of the flow result endpoint, which can eliminate the influence of small flow errors. 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. In addition, robust statistics and the accuracy measures derived from the Middlebury benchmark are employed to evaluate the robustness of optical flow in the proposed method.

Introduction

It is undisputed that optical flow estimation has been the one of the most interesting research directions in the field of computer vision, image processing, and pattern recognition. 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. Additionally, optical flow evaluation theory is one of the most important factors in the development and improvement of optical flow estimation[1–7].

The purpose of optical flow evaluation is to expose the defects of current optical flow methods in order to promote the development of optical flow estimation research, which focuses on the more difficult problems of optical flow estimation. Specifically, the goal of optical flow evaluation is to determine what phenomenon causes the hard task in optical flow computing by providing the ground truth of the test image sequence set contained the location of the movement boundary and low-texture area information. we present problems with traditional evaluation measures in current

Copyright © 2017, the Authors. Published by Atlantis Press.
This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

evaluation methodology and the proposed improved evaluation methodology with corrected evaluation measures is described.

To provide a convincing evaluation methodology, Baker [7] proposed an extended set of performance measures to focus attention on current optical flow algorithmic problems. It is commendable that the Middlebury benchmark was the acknowledged evaluation benchmark for optical flow estimation. Notwithstanding that the Middlebury benchmark had been the representative optical flow evaluation benchmark, the performance improvement of current optical flow algorithms required more challenging test image sequences. To motivate work on the next generation optical flow algorithms in overcoming more difficult tasks, Butler [8] introduced a naturalistic database for optical flow evaluation derived from the open-source CGI movie Sintel. In contrast to the Middlebury database, this MPI database [9] contains longer and more varied test image sequences with image degradations such as motion blurs, defocus blurs, and atmospheric effects.

For problems in current optical flow evaluation methodology, this paper proposes an improved methodology by correcting the traditional evaluation measures of optical flow. This paper contains the following content: (1) we present problems with traditional evaluation measures in current evaluation methodology and the proposed improved evaluation methodology with corrected evaluation measures is described. (2) Some experimental results are presented and a brief summary is concluded.

Current and Our Improved Evaluation Methodology

Current Evaluation Methodology

The current Middlebury evaluation methodology for optical flow was proposed by Baker, and contained mainly three components: error measures, error statistics, and region masks.

The current computational formula of the endpoint error can be expressed as follows:

$$EE = \sqrt{(u - u_{GT})^2 + (v - v_{GT})^2} \tag{1}$$

To provide a more equitable measurement, a normalized magnitude of the vector difference between the correct and estimated flow vectors was proposed by McCane [6]:

$$E_M = \begin{cases} \dfrac{\|\mathbf{v}_{GT} - \mathbf{v}_E\|}{\|\mathbf{v}_{GT}\|} & \text{if } \|\mathbf{v}_{GT}\| \geq T, \\ \left| \dfrac{\|\mathbf{v}_E\| - T}{T} \right| & \text{if } \|\mathbf{v}_{GT}\| < T \text{ and } \|\mathbf{v}_E\| \geq T, \\ 0 & \text{if } \|\mathbf{v}_{GT}\| < T \text{ and } \|\mathbf{v}_E\| < T, \end{cases} \tag{2}$$

where E_M is the error measure, $\mathbf{v}_{GT}$ denotes the correct flow vector, $\mathbf{v}_E$ denotes the estimated flow vector, and $T = 0.5$ pixel is the threshold.

Our Improved Evaluation Methodology

Although the current evaluation methodology has been used widely in optical flow, it still contains the ineffective measures described above. Now a corrected evaluation methodology for optical flow is presented which may be better suited to the evaluation of modern optical flow algorithms.

Although the angle error can indicate the performance of the optical flow, it has limitations in some cases. For example, as can be seen in Fig. 1, there are two computed optical flow vectors (V_1,V_2) and corresponding ground truth vectors (V_{GT1},V_{GT2}). In this case, the larger optical flow vector v_1 has the larger endpoint error even though the computed vectors (V_1,V_2) have the same angle errors.

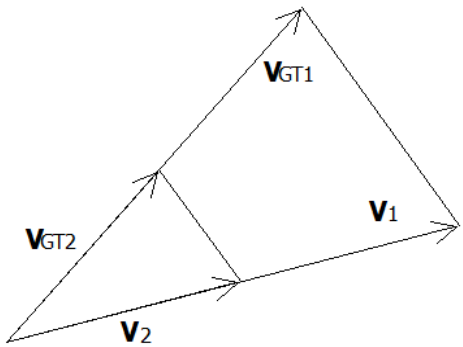

Fig.1 The expression of the example for the endpoint error

To evaluate the optical flow comprehensively, the error measure of the endpoint is indispensable. The traditional error measure of the endpoint is the magnitude of the vector difference between the computed optical flow and the ground truth, which cannot indicate the true magnitude difference between the computed optical flow and the ground truth. To obtain a precise error measure of the endpoint, the corrected measure for the magnitude error of the optical flow is presented as below:

$$EE = \left\| \mathbf{v} \right| - \left| \mathbf{v}_{\mathrm{GT}} \right\| \tag{3}$$

where V denotes the computed optical flow and V_{GT} denotes the ground truth. According to the corrected error measure of the endpoint, the endpoint error is a scalar which only indicates the length difference between the computed optical flow and the ground truth.

The proposed error measure of the endpoint can indicate the true length difference between the computed optical flow and the ground truth; however, it may still portray a seemingly small endpoint error for small optical flows that may actually be relatively larger. To show the performance of the optical flow equitably, we augment the normalization error measure of the endpoint to supplement the error measurement:

$$NEE = \frac{EE}{|\mathbf{v}_{GT}|} = \frac{\left||\mathbf{v}| - |\mathbf{v}_{GT}|\right|}{|\mathbf{v}_{GT}|} \tag{4}$$

where V denotes the computed optical flow and V_{GT} denotes the ground truth. With the proposed normalization error measure of the endpoint, the problem of endpoint error for small optical flows being overlooked owing to domination by the endpoint error of large optical flows could be eliminated, providing an objective measurement result for optical flows.

Experiments and Analysis

For the comprehensive evaluation of optical flow, the endpoint error is an indispensable error measurement which indicates the magnitude of the distance between the estimated flow and the ground truth. To show the differences in endpoint error between the proposed benchmark and the Middlebury benchmark, the statistical results of the endpoint error in the evaluation methods with the Middlebury and proposed benchmarks are shown in Tables 1 and 2, respectively.

Table.1 EE with Middlebury benchmark

METHOD	ALL	Dimetrodon	Grove3	Rubber whale	Urban2
LDOF	Avg	0.118	0.489	0.115	0.256
	STD	0.126	1.179	0.310	1.167
	R0.5	0.977	0.814	0.955	0.941
	R1.0	1.000	0.887	0.977	0.965
	R2.0	1.000	0.942	0.991	0.977
	A50	4.286	5.012	2.117	3.964
	A75	8.330	17.954	3.870	7.659
	A95	22.48	147.38	25.983	36.314
GLC-TV	Avg	0.179	0.510	0.120	0.345
	STD	0.170	1.263	0.364	1.454
	R0.5	0.935	0.819	0.964	0.908
	R1.0	0.999	0.884	0.981	0.952
	R2.0	1.000	0.937	0.989	0.972
	A50	7.000	6.154	2.743	5.469
	A75	13.42	16.330	4.685	10.932
	A95	31.21	156.61	18.973	54.815
SODOF	Avg	0.202	0.591	0.136	0.288
	STD	0.180	1.134	0.295	0.587
	R0.5	0.933	0.733	0.962	0.879
	R1.0	0.995	0.867	0.981	0.960
	R2.0	1.000	0.936	0.995	0.982
	A50	8.717	12.295	4.132	7.982
	A75	14.86	31.309	6.819	16.837
	A95	30.98	148.71	23.502	49.956
Correlation Flow	Avg	0.234	0.362	0.081	0.280
	STD	0.248	0.876	0.173	0.718
	R0.5	0.887	0.875	0.990	0.894
	R1.0	0.975	0.946	0.995	0.966
	R2.0	1.000	0.967	0.998	0.985
	A50	8.458	8.076	3.076	7.681
	A75	16.86	16.940	4.667	15.371
	A95	47.64	62.599	11.259	45.403
Classic+nl	Avg	0.117	0.399	0.079	0.155
	STD	0.115	1.099	0.265	0.494
	R0.5	0.985	0.868	0.981	0.964
	R1.0	0.999	0.924	0.988	0.983
	R2.0	1.000	0.953	0.995	0.992
	A50	4.921	4.646	1.989	3.987
	A75	8.604	11.504	3.157	7.286
	A95	18.15	105.06	10.574	22.926

Table.2 EE with the proposed benchmark

METHOD	ALL	Dimetrodon	Grove3	Rubber whale	Urban2
LDOF	EX	0.082	0.356	0.062	0.207
	EXSTD	0.108	0.899	0.162	0.944
	R0.1	0.738	0.609	0.875	0.742
	R0.3	0.941	0.775	0.957	0.916
	R0.5	0.991	0.839	0.976	0.946
	A95	0.324	1.715	0.262	0.560
GLC-TV	EX	0.111	0.341	0.062	0.392
	EXSTD	0.137	0.907	0.200	1.125
	R0.1	0.618	0.607	0.895	0.501
	R0.3	0.930	0.808	0.972	0.794
	R0.5	0.973	0.861	0.983	0.880
	A95	0.344	1.695	0.189	1.607
SODOF	EX	0.136	0.414	0.084	0.232
	EXSTD	0.144	0.818	0.207	0.536
	R0.1	0.498	0.390	0.802	0.515
	R0.3	0.898	0.689	0.954	0.804
	R0.5	0.971	0.788	0.977	0.900
	A95	0.441	1.806	0.280	0.784
Correlation Flow	EX	0.123	0.230	0.049	0.213
	EXSTD	0.165	0.602	0.098	0.667
	R0.1	0.607	0.556	0.907	0.533
	R0.3	0.902	0.845	0.988	0.840
	R0.5	0.965	0.916	0.995	0.922
	A95	0.415	0.816	0.141	0.665
Classic+nl	EX	0.072	0.260	0.044	0.119
	EXSTD	0.099	0.746	0.153	0.422
	R0.1	0.760	0.684	0.933	0.747
	R0.3	0.969	0.845	0.980	0.935
	R0.5	0.992	0.890	0.990	0.969
	A95	0.237	1.264	0.129	0.361

For a visual comparison, the best result of each statistics item for one sequence is highlighted with red and in bold text in Tables 1 and 2. Note that the distribution of the best results in Table 2 is
partially changed compared with Table 1, which indicates that the proposed measure of endpoint error has different effectiveness compared with the current measure. For example, in the statistical
results of the Dimetrodon sequence, the LDOF method won ten best results in the statistics
measurements with the Middlebury benchmark; however, it only won two best results in the statistics measurements using the proposed benchmark. The given example is not meant to emphasize the performance ranking of the LDOF method, but to show the differences in the measures of endpoint error between the proposed evaluation benchmark and the Middlebury benchmark.

In Table 2, the numeric values of the statistics results are much smaller than the results in Table 1 since the proposed measure of endpoint error is the pure difference of the magnitude of the distance between the estimated flow and the ground truth, which avoids the detrimental influence of the angle deviation of the flow results. For general evaluation, the statistics results of the EX and EXSTD indicate the holistic performance of the flow results, where EX reflects the average error and the EXSTD reflects the deviation. For example, in the statistics results of the Dimetrodon sequence, the Classical+nl method won the best result for EX with the Disc mask; however, the corresponding statistics result of EXSTD for this method is large, which indicates that the endpoint error of the Classical+nl method corresponding to the Disc mask of the Dimetrodon sequence follows a discrete distribution.

For the robust statistics index R*X* and the accuracy measure index A*X*, the statistical results of R0.1, R0.3, and R0.5 show the robustness of the flow results and the statistical results of A95 show the robust convergence of the accuracy of the flow results. Similar to the Middlebury benchmark, the proposed robust statistics of the endpoint error provide extensional evaluation for the performance of flow results which can reflect their robust convergence. For the analysis of the endpoint errors in different image regions, the statistics results of the endpoint error with the All mask show the performances of the evaluation method in different image regions; the statistical results in Table 2 indicate that the Disc mask is an undisputed challenge for optical flow.

The statistics results in Table 2 show that the proposed measure of the endpoint error has a different character compared with the current measure of the Middlebury benchmark since the former indicates the absolute difference of the magnitude of the flow result; this should be the genuine endpoint error in an optical flow.

Conclusions

In this paper, for the problem of objective evaluation, this paper proposed an improved evaluation benchmark by modifying the metrics and statistics of optical flow error. Major contributions included the following. For metrics of the flow error, we corrected measures of endpoint error of the flow result, and the corrected measures

were found to reflect actual performance of the evaluated flow. In addition, a measure of the normalization endpoint error was employed to indicate the relative error of the endpoint, which can avoid the problem of vanishing small flow errors. 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.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (61401190, 61462062), the Key Project of the Natural Science Foundation of Jiangxi Province (20133ACB20004), the Advantage Subject Team Project of Jiangxi Province(20152BCB24004), the Aeronautical Science Foundation of China (2015ZC56009), the Foundation of Key Laboratory of Nondestructive Testing, Ministry of Education(ZD201529001) and the Foundation of Key Laboratory of Jiangxi Province for image processing and pattern recognition(TX201604001).

References

[1] Verri A, Poggio T. Motion field and optical flow: Qualitative properties[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1989, 11(5): 490-498.

[2] 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.

[3] Otte M, Nagel H H. Estimation of optical flow based on higher-order spatiotemporal derivatives in interlaced and non-interlaced image sequences[J]. Artificial Intelligence, 1995, 78(1): 5-43.

[4] Baraldi P, Sarti A, Lamberti C, et al. Evaluation of differential optical flow techniques on synthesized echo images[J]. IEEE Transactions on Biomedical Engineering, 1996, 43(3): 259-272.

[5] Galvin B, McCane B, Novins K, et al. Recovering Motion Fields: An Evaluation of Eight Optical Flow Algorithms[C]. The British Machine Vision Conference, 1998: 195-204.

[6] McCane B, Novins K, Crannitch D, et al. On benchmarking optical flow[J]. Computer Vision and Image Understanding, 2001, 84(1): 126-143.

[7] 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.

[8] Butler D J, Wulff J, Stanley G B, et al. A naturalistic open source movie for optical flow evaluation[C]. European Conference on Computer Vision, 2012: 611-625.

[9] http://sintel.is.tue.mpg.de/