Confidence Measure for Block-Based Motion Vector Field

Karen Simonyan, Sergey Grishin, and Dmitriy Vatolin
Graphics and Media Laboratory, Department of Computational Mathematics and Cybernetics
Moscow State University, Moscow, Russia
{simonyan, sgrishin, dmitriy}@graphics.cs.msu.ru

Abstract

In this paper we propose a confidence measure for block-based motion vector field. The measure is calculated as an average of two *a posteriori* estimates which reflect various aspects of MVF accuracy: motion-compensated interframe difference distribution and motion vectors distribution. Experimental results show that the proposed measure outperforms its contemporary counterpart while demanding less information about the motion estimation process.

Keywords: *confidence measure, motion estimation, motion vector field.*

1. INTRODUCTION

Motion information is used in most contemporary video processing algorithms, as it allows getting benefit from video redundancy, thereby enhancing the algorithm performance. Since ground-truth Motion Vector Field (MVF) is usually not available, a Motion Estimation (ME) algorithm is applied to calculate the motion between video frames. The correspondence of the calculated motion to a ground-truth motion is one of the key issues, as the utilization of wrong motion information (*e.g.* caused by aperture problem or occlusion [7]) can lead to artifacts in the areas of the processed video, where this information is used. Therefore, certain objective criterion is needed to express the correspondence.

The confidence measure is such a criterion. It can be treated as a probability that an estimated Motion Vector (MV) is equal to a ground-truth MV. The confidence measure is a universal means to control MVF correctness, as it can be applied in two scenarios:

- the measure can be incorporated directly into the conventional ME algorithm to detect wrongly estimated MVs. A special postprocessing is then applied to these MVs to improve their accuracy;
- the measure can be a part of video postprocessing algorithm (*e.g.* frame rate up-conversion or deinterlacing) executed on the decoder side, where the information about the ME process is inaccessible. In this case a separate branch of the algorithm can be provided to handle the processing of areas, for which no reliable motion information is available.

In this paper we propose a confidence measure for block-based MVF, which takes different spatial and temporal cues into account. The results of the comparison with the method proposed by Patras *et al.* [1] justify the superiority of the proposed method.

The rest of the paper is organized as follows. In Section 2 a review of the related work is given. In Section 3 the proposed algorithm is described. Experimental results are presented in Section 4. Section 5 concludes the paper.

2. RELATED WORK

There are two major approaches [1] to a confidence measure calculation. The first is to estimate *a priori* confidence to a MV before its explicit calculation. Generally, such methods take certain spatial cues into account, *e.g.* spatial luminance derivatives [3], [4]. The basic idea is to determine the areas where the aperture problem can arise. This information can further be used to decide whether ME can be confidently applied to certain area, or not [2]. The major drawback of this technique is a small application field, as the approach is almost useless when a confidence to an estimated MVF is to be obtained.

The second approach assumes the calculation of *a posteriori* confidence to already estimated MVF. In this case not only the spatial cues can be used, but proper MVF modeling and analysis as well. The methods following this approach can be divided into two groups according to the MVF structure. The first group is formed by the methods intended for optical flow confidence estimation [5], [6]. The second group consists of the algorithms estimating confidence to a block MVF, *e.g.* the MVF calculated by block matching ME. These algorithms are of particular interest, since block matching ME is widely used in conventional video processing systems. In [7] To *et al.* proposed a confidence measure based on frame phase information, thus limiting the application field of the method within cases where phase correlation ME is used. Lundmark *et al.* [8] used the weighted sum of Motion-Compensated Interframe Difference (MCID) to obtain the confidence value. This algorithm is applicable in the case of occlusion, but in the case of aperture problem it fails.

Recently Patras *et al.* [1] introduced a confidence measure in the probabilistic framework. They proved that the block-based ME minimizing the Sum of Absolute Differences (SAD) is equivalent to a maximum likelihood estimator of MVs, assuming that the MCID follows the Laplacian distribution. Considering candidate motion vectors for each block to be known, *a posteriori* probability of calculated MV being equal to a ground-truth MV is estimated. However, the dependence on the candidate set is a disadvantage, since:

- all the candidates must be known. Therefore, the measure can be calculated only while performing ME. This fact impedes the application of the measure in video processing on the decoder side, as candidate motion vectors are unavailable in a video stream;
- contemporary block matching methods [9] use various MVF consistency cues to reduce the candidate set, thus gaining efficiency. However, this can lead to erroneous confidence estimates. For instance, ME algorithm can wrongly construct a candidate set of one MV, nevertheless Patras *et al.* measure will assign a unity (*i.e.* highest) confidence to the vector because there are no other vectors in the set. Thereby, the measure can be

applied only to pattern search ME methods, *e.g.* full-search.

3. PROPOSED CONFIDENCE MEASURE

The proposed confidence measure is derived in the following way. Two *a posteriori* estimates of MV confidence are obtained. The first estimate, P_{MCID}, is based on MCID analysis. The second estimate, P_{MVF}, takes MVF distribution into account. These estimates are combined, resulting in a confidence measure. Various schemes of estimates combination exist. Kittler *et al.* [10] argued that in practice simple averaging often produces better results than more sound techniques do. From our experiments, we came to the same conclusion. So, the measure is calculated as:

$$P = \frac{P_{MCID} + P_{MVF}}{2}. \tag{1}$$

The derivation of the estimates P_{MCID} and P_{MVF} is given in Sections 3.1 and 3.2 respectively.

3.1 MCID analysis

As in [1], we consider the case where SAD is used as an error function for MV and assume that the MCID follows the Laplacian distribution with a zero mean:

$$P(I(x)) = \frac{1}{2\lambda_B}\exp\left(-\frac{|I(x)|}{\lambda_B}\right), \tag{2}$$

where $I(x)$ is a MCID value of pixel x from block B, λ_B is a parameter, specific to the block. The variance σ_B^{MCID} of block MCID is linked with λ_B by the following expression:

$$\sigma_B^{MCID} = \sqrt{2}\lambda_B. \tag{3}$$

Given the sample $\{I(x) | x \in B\}$ a maximum likelihood estimate of λ_B can be derived as follows:

$$\begin{aligned}\lambda_B &= \arg\max_{\lambda_B} \prod_{x\in B} P(I(x)) = \\ &= \arg\max_{\lambda_B}\left(\left(\frac{1}{2\lambda_B}\right)^{|B|}\cdot\exp\left(-\frac{\sum_{x\in B}|I(x)|}{\lambda_B}\right)\right) = \\ &= \arg\max_{\lambda_B}\left(\left(\frac{1}{2\lambda_B}\right)^{|B|}\cdot\exp\left(-\frac{\mathrm{SAD}(B)}{\lambda_B}\right)\right) = \\ &= \arg\max_{\lambda_B}\left(|B|\cdot\ln\frac{1}{2\lambda_B} - \frac{\mathrm{SAD}(B)}{\lambda_B}\right) = \frac{\mathrm{SAD}(B)}{|B|}.\end{aligned} \tag{4}$$

Thus, only the SAD value of the block MV is needed to obtain the estimate of λ_B.

In [1] it was suggested that σ_B^{MCID} depends on block luminance variance σ_B^I: the larger σ_B^I, the larger σ_B^{MCID}. The argumentation is as follows. If the value of σ_B^I is low (uniform block), and a good reference block was found by ME, the variance of motion-compensated difference between the current and the reference blocks σ_B^{MCID} will also be low. On the contrary, it is quite natural to suppose that for large σ_B^I values (*e.g.* block containing fine texture) the value of σ_B^{MCID} will be large as no perfect match is usually possible in such cases. We approximate this dependence with a linear function:

$$\sigma_B^{MCID} = \alpha_0 + \beta_0 \cdot \sigma_B^I, \tag{5}$$

where α_0 and β_0 are constants specific for each video frame.

The dependence between σ_B^I and σ_B^{MCID}, plotted for one frame of *Fly* test video (test videos are described in Section 4) is presented in Figure 1 together with the approximant function.

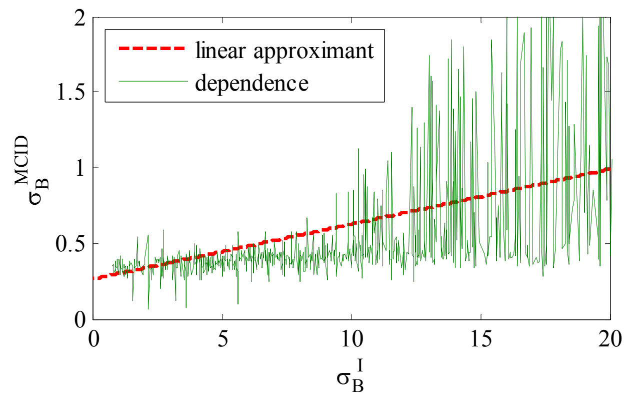

Figure 1: The dependence of σ_B^{MCID} on σ_B^I for *Fly* frame and the correspondent linear approximant.

Substituting (3) and (4) into (5), we derive:

$$\frac{\mathrm{SAD}(B)}{|B|} = \alpha + \beta \cdot \sigma_B^I, \tag{6}$$

where $(\alpha, \beta)^T \equiv \frac{1}{\sqrt{2}}(\alpha_0, \beta_0)^T$.

Assuming that $\mathrm{SAD}(B)$ and σ_B^I are known for each block B of the frame, α and β are calculated using linear least squares method, which leads to the following estimates:

$$\begin{aligned}\alpha &= \frac{\sum_B\left(\sigma_B^I\right)^2 \cdot \sum_B \frac{\mathrm{SAD}(B)}{|B|} - \sum_B \sigma_B^I \cdot \sum_B \sigma_B^I \frac{\mathrm{SAD}(B)}{|B|}}{n\sum_B\left(\sigma_B^I\right)^2 - \left(\sum_B \sigma_B^I\right)^2}, \\ \beta &= \frac{n\sum_B \sigma_B^I \frac{\mathrm{SAD}(B)}{|B|} - \sum_B \sigma_B^I \cdot \sum_B \frac{\mathrm{SAD}(B)}{|B|}}{n\sum_B\left(\sigma_B^I\right)^2 - \left(\sum_B \sigma_B^I\right)^2},\end{aligned} \tag{7}$$

where n is the number of blocks in a frame.

To obtain a P_{MCID} value for a block B, an estimate of its SAD value is calculated:

$$\text{SAD}'(B) = |B|\left(\alpha + \beta \cdot \sigma_B^I\right), \tag{8}$$

and finally,

$$P_{MCID} = \min\left(\max\left(\frac{\text{SAD}'(B)}{\text{SAD}(B)}, 0\right), 1\right). \tag{9}$$

3.2 MVF analysis

The estimate P_{MVF} reflects the confidence to a MV based on the MVF distribution analysis. Here we utilize the following heuristics. If the block MV is close in some sense to the MVs of adjacent blocks, it is likely that the blocks correspond to a part of object, exhibiting uniform motion, so the confidence is assigned to the block depending on the size of the object part. On the contrary, if the block MV differs from adjacent blocks' MVs, it can be a wrong vector. This cue leads to the following algorithm:

- cluster the MVF in some manner;
- let C_B be the cluster which the block B belongs to; then P_{MVF} is calculated as:

$$P_{MVF} = \min\left(\frac{|C_B|}{thresh}, 1\right), \tag{10}$$

where $thresh$ is a threshold, which value is chosen empirically. In our experiments the value was set to 0.5% of the frame area.

As it can be seen, the performance of this part of confidence estimation algorithm essentially depends on the clustering algorithm chosen. We employed a modification of an agglomerative hierarchical clustering algorithm [11] that maintains clusters' spatial consistency. Such choice of the clustering algorithm (that will be discussed further) was driven by several reasons:

- the number of clusters in a frame is *a priori* unknown;
- the clusters on the output of the algorithm must be spatially consistent frame regions.

The distance between blocks is introduced as the L_2 distance between corresponding MVs:

$$d(B_1, B_2) = \|MV(B_1) - MV(B_2)\|_2, \tag{11}$$

where $MV(B_{1,2})$ is the MV of block $B_{1,2}$. The distance between clusters C_1 and C_2 is calculated as the mean distance between cluster elements (the so-called average linkage clustering):

$$d(C_1, C_2) = \frac{1}{|C_1| \cdot |C_2|} \sum_{B_i \in C_1} \sum_{B_j \in C_2} d(B_i, B_j). \tag{12}$$

Unlike conventional agglomerative hierarchical clustering, the proposed clustering algorithm at each step merges *spatially adjacent* clusters, distance between which is minimal. The merging process is stopped when the minimal distance between adjacent clusters exceeds a certain threshold. Empirically the threshold was set to the mean length of the frame MVs; thereby the clustering adapts to the motion activity in the frame. An example of MVF clustering for one frame of *Fly* test video is demonstrated in Figure 2; different clusters are painted with different colors.

Figure 2: Clustered MVF of *Fly* frame.

4. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed measure, we used a test set of two synthetically rendered video sequences, *Fly* and *Tower*, for which ground-truth MVF is available. These videos are in 720×576 resolution and consist of 151 frames. Full-search ME algorithm (16×16 blocks were used, MVs were estimated with a quarter-pixel accuracy) was applied to these videos, resulting in MVFs that were further used to compare confidence measures. After that, two confidence estimation algorithms, the proposed one and the method by Patras *et al.* [1], were used to obtain the confidence measure values for these MVFs. Then, Spearman and Kendall rank correlation coefficients [12] were calculated between each of these measures and the ground-truth error score GT_Δ. This score is defined as:

$$GT_\Delta = \|MV_{FS}(B) - MV_{GT}(B)\|_2, \tag{13}$$

where $MV_{FS}(B)$ is a MV of block B, obtained by the full-search ME, $MV_{GT}(B)$ is a ground-truth MV of this block.

The so-called Spearman's rho ρ and Kendall's tau τ are quite popular measures to evaluate the correspondence between different scores for the same group of objects. Thus, it can be determined how much the compared confidence measures correspond to the ground-truth error.

Table I
Comparison of median ρ and τ for the compared algorithms

Confidence measure algorithm	Video sequence			
	Fly		Tower	
	median ρ	median τ	median ρ	median τ
Proposed	0.3219	0.2867	0.5628	0.4631
Patras *et al.*	0.2724	0.2388	0.4596	0.3491

The results of the comparison of ρ and τ median values calculated for test videos are given in Table I. Plots of Spearman's rho for *Fly* and *Tower* videos are presented in Figure 3 and Figure 4 respectively. The plots of Kendall's tau are not given as the results of τ comparison very closely follow those of ρ comparison. As it can be seen, the proposed confidence measure outperforms Patras *et al.* measure, providing acceptable rank correlation with the ground-truth score.

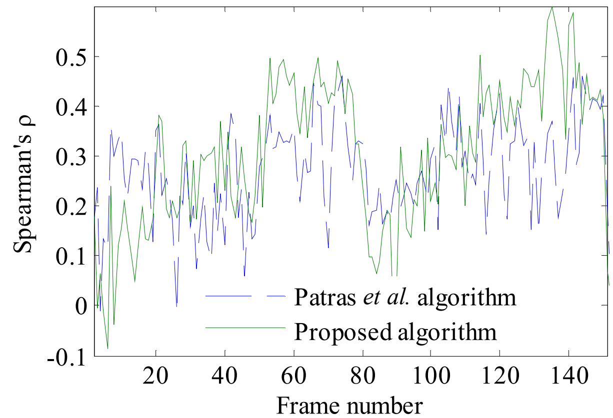

Figure 3: Comparison of Spearman's rho on *Fly* video.

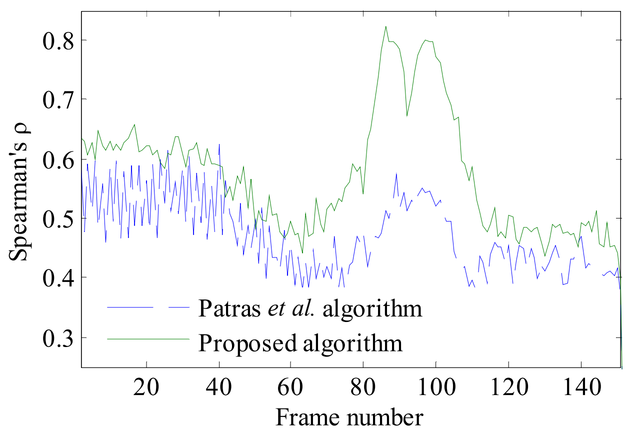

Figure 4: Comparison of Spearman's rho on *Tower* video.

5. CONCLUSION

In this paper we presented a new confidence measure estimation algorithm. It was argued that the algorithm produces plausible confidence estimates in terms of rank correlation with the ground-truth error score. At the same time, only the current video frame and its motion vectors with correspondent errors are needed to calculate the measure; no supplementary information is needed.

The performance of the algorithm can be further improved by employing more advanced clustering and robust approximation of the dependence between the variance of block intensity and the variance of its motion-compensated interframe difference. Moreover, motion trajectory smoothness can be an additional cue to be utilized in the confidence estimate derivation.

6. REFERENCES

[1] I. Patras, E. A. Hendriks, and R. L. Lagendijk. Probabilistic Confidence Measures for Block Matching Motion Estimation. *IEEE Trans. on CSVT*, Aug. 2007, vol.17, pp. 988–995.

[2] P. Sand and L. McMillan. Efficient selection of image patches with high motion confidence. In *Proc. IEEE ICIP*, Rochester, NY, 2002, pp. 293–296.

[3] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. In *Proc. IJCV*, 1994, vol. 12, no. 1, pp. 43–77.

[4] T. Yoshida, H. Katoh, and Y. Sakai. Block matching motion estimation using block integration based on reliability metric. In *Proc. IEEE ICIP*, 1997, vol. 2, pp. 152–155.

[5] E. P. Simoncelli, E. H. Adelson, and D. J. Heeger. Probability distributions of optical flow. In *Proc. IEEE ICCVPR*, Maui, HI, Jun. 1991, pp. 310–315.

[6] A. Dev, B. J. A. Krose, and F. C. A. Groen. Confidence measures for image motion estimation. In *Proc. RWC Symposium*, 1997, pp. 199–206.

[7] L. To, M. Pickering, M. Frater, and J. Arnold. A motion confidence measure from phase information. In *Proc. IEEE ICIP*, Oct. 2004, vol. IV, pp. 2583–2586.

[8] A. Lundmark, H. Li, and R. Forchheimer. Motion vector certainty reduces bit rate in backward motion estimation video coding. In *Proc. SPIE VCIP*, Jun. 2000, pp. 95–105.

[9] Ahmad, I., Zheng, W.G., Luo, J.C., Liou, M. A Fast Adaptive Motion Estimation Algorithm. *IEEE Trans. on Image Processing*, 2006, vol. 16, issue 3, pp. 420–438.

[10] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. On combining classifiers. *IEEE Trans. on PAMI*, vol. 20, no. 3, pp. 226–239, Mar. 2000.

[11] R. Duda, P.E. Hart, D.G. Stork. *Pattern Classification*. 2nd ed., Wiley-Interscience, 2001.

[12] D. Sheskin. *Handbook of parametric and nonparametric statistical procedures*. 2nd ed., Chapman & Hall/CRC, 2000.

Acknowledgement

The authors would like to thank the Ph.D. student Anatoly Zapadinsky for synthetic video sequences and corresponding ground-truth MVF.

About the authors

Karen Simonyan is a student at Graphics and Media Laboratory, Department of Computational Mathematics and Cybernetics, Moscow State University. His research interests include motion estimation, super-resolution, frame-rate conversion and adjacent field. His contact email is simonyan@graphics.cs.msu.ru

Sergey Grishin is a Ph.D. student at Graphics and Media Laboratory, Department of Computational Mathematics and Cybernetics, Moscow State University. His research interests include deblocking, motion estimation, frame-rate conversion, stereo vision and video segmentation. His contact email is sgrishin@graphics.cs.msu.ru

Dmitriy Vatolin is an expert in image, video and data compression with more than 14 years experience; Ph.D. in image compression, IEEE Member. He is a senior researcher and the head of the Video Group at Graphics and Media Laboratory, Department of Computational Mathematics and Cybernetics, Moscow State University. His contact email is dmitriy@graphics.cs.msu.ru