A NUMERICAL COMPARISON OF COMPRESSED DOMAIN APPROXIMATIONS TO OPTICAL FLOW

Miguel Coimbra, Mike Davies

Queen Mary, University of London
Department of Electronic Engineering
miguel.coimbra@elec.qmul.ac.uk, mike.davies@elec.qmul.ac.uk

ABSTRACT

Optical flow is an invaluable tool for numerous video processing tasks. However, the high computational cost of optical flow estimation algorithms is a serious limitation for practical systems. A fast alternative exists in digital video encoding standards that offer an approximation to the optical flow field in the form of motion vectors. Additional filtering and processing of these vectors can improve this approximation but most authors choose these methods empirically without accurate knowledge of its full effects. This paper analyses and compares several approximations present in literature both visually and numerically showing the importance of both spatial median filtering and confidence maps. Experiments are done using real world data obtained from underground train surveillance cameras.

1. INTRODUCTION

The rising popularity of digital video is leading to an increase in the research of algorithms using *compressed domain information*. This information is defined as data obtained from a decoding stage of a digital video encoding standard, before the fully decoded image is obtained. Although a variety of standards exists (MPEG-1,2,4; H-26x), most of them share core technologies such as block-based DCT and motion estimation. This information has been used in the past for tasks such as video indexing [1], camera motion estimation [2] and motion segmentation [3]. Motion vectors obtained from MPEG video are used as an approximation to optical flow by these authors after a normalization step [1], median filtering [2] or more complex filtering [3]. Comparisons of the quality of the approximations are difficult since they are used for different purposes thus only subjective conclusions can be drawn.
Following work by Barron et al. [4] and Galvin et al. [5] that analytically compared several optical flow estimation algorithms, different compressed domain approximations will be tested and compared. The previous authors used synthetic sequences where ground truth is readily available and estimation accuracy can be numerically evaluated. In this paper, the accuracy of the approximation to optical flow will be obtained instead by direct comparison with the popular Lucas and Kanade method [6]. Test sequences can thus be obtained from real underground station surveillance cameras, reflecting both the importance of these approximations to real world applications as well as the necessity of very fast processing of automated surveillance systems.
Section 2 will describe the various approximations used as well as the numerical evaluation measures. Experimental results will be shown in section 3 and final conclusions drawn on section 4.

2. METHODS

Different compressed domain approximations can be obtained from the MPEG-2 standard [7] by usage of DCT coefficients and motion vectors. Knowledge of this standard is important for correct understanding of the methods and experiments presented here.

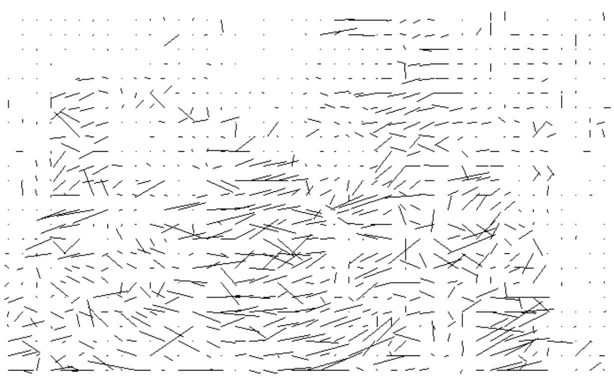

Figure 1 – Raw motion vectors

The motion field obtained by using MPEG-2 motion vectors (*raw motion vectors*) can be seen on Fig. 1. Its corresponding decoded image is in Fig. 2. It can be observed that *raw motion vectors* are very noisy and only a vague approximation to the scene's optical flow.

Figure 2 – Decoded image

Four techniques are typically used to increase the quality of this approximation: normalization rules, median filtering, increased resolution and confidence thresholding. These techniques are independent and either one or all four can be applied to the raw motion vectors.

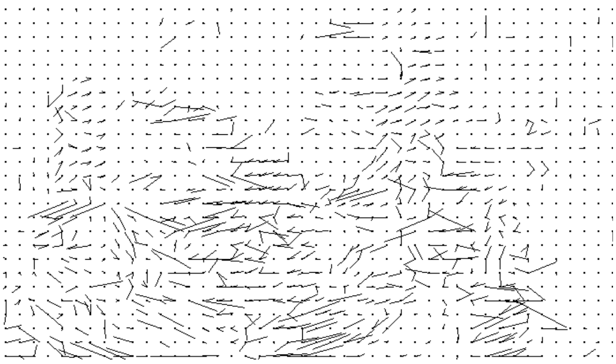

Figure 3 – Normalized motion vectors

Normalization rules – Not every MPEG-2 *macroblock* has one motion vector. Some have none (I blocks) and some have two (*interpolated blocks* in B-frames) therefore some normalization rules are used by authors to obtain one vector per macroblock. These rules make the motion field consistent and independent of the MPEG-2 picture type. It is difficult to choose a standard set of rules since they are mostly ad-hoc so, for comparison purposes, the rules explained in [8] are used (results seen on Fig. 3):

- Macroblocks with no motion vector have the same movement as in the previous image.
- When a macroblock has two motion vectors, the one pointing back is reversed and added to the one pointing forward.
- Macroblocks with one motion vector in B-Pictures are scaled (since motion vectors in P-Pictures typically span three images, therefore are larger).
- Skipped macroblocks in I-pictures have no movement while in P-Pictures have movement similar to the previous block.

Spatial median filtering – This type of filter is very effective at removing *spot noise* from a signal while minimizing distortion. As such, random noise vectors are mostly eliminated by the usage of two spatial 3x3 median filters: one for the horizontal and another for the vertical direction of vectors. Results can be seen on Fig. 4.

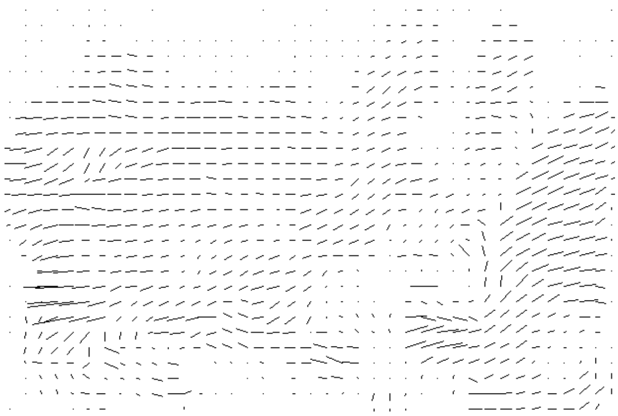

Figure 4 – Median filtered, normalized motion vectors

Increased resolution – Resolution can be artificially increased by using the same 16x16 vector for all four corresponding 8x8 blocks. This step makes more sense when combined with *confidence thresholding* since *confidence maps* have 8x8 resolution.

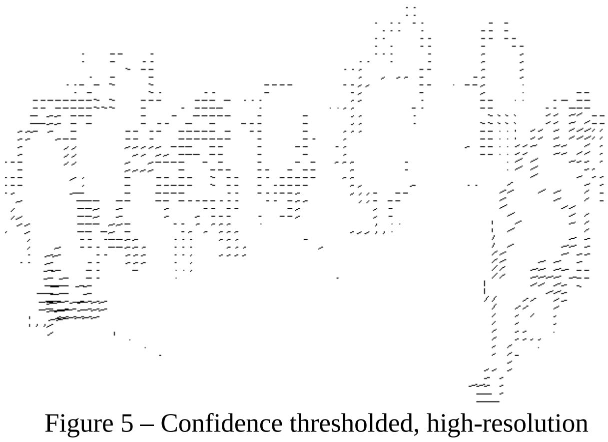

Figure 5 – Confidence thresholded, high-resolution motion field.

Confidence thresholding – It has previously been shown in [8] how *confidence maps* can be obtained from selected DCT coefficients. Using update rules, these maps are defined for all MPEG-2 frames regardless of their *type*. These maps reflect the spatial luminance gradient strength for each 8x8 block thus signaling situations where vector inaccuracy is caused by *aperture problems*. Fig. 5 shows the improvement over Fig. 4 by using both *confidence thresholding* and *increased resolution*.

Figure 6 – Sub-sampled Lucas Kanade motion field

Although visual inspection is important, some form of numerical comparison is required for objective comparison of the different approximations to the LK (Lucas Kanade) method. Two different analytical measures were chosen for the numerical evaluation of the several algorithms:

Mean square error (MSE) - of the vector magnitude. This measure is defined in equation (1) where $\overrightarrow{LK}$ is the Lucas Kanade field, $\overrightarrow{OF}$ is the compressed domain approximation to optical flow and X and Y are the width and height of the vector matrix.

$$MSE = \frac{1}{XY}\sum_{x,y}\left(\left|\overrightarrow{LK}(x,y)\right| - \left|\overrightarrow{OF}(x,y)\right|\right)^2 \qquad (1)$$

This measure reflects the energy of the noise present. Generally, it gives us an indication of how many and how strong the noise vectors are.

Mean angular error (MAE) - of the vector field. This measure is defined in equation (2) where N is the number of field locations where $\overrightarrow{LK}$ and $\overrightarrow{OF}$ are non-zero.

$$MAE = \frac{1}{N}\sum_{x,y}\left|\angle\overrightarrow{LK}(x,y) - \angle\overrightarrow{OF}(x,y)\right| \qquad (2)$$

This measure reflects the accuracy of the direction of the optical flow approximation. Its limitation is that it is not defined for all field locations.

These measures complement each other and should be analyzed together. MSE gives us an estimation of how many incorrect vectors exist and MAE, while ignoring this noise, measures the actual accuracy of the estimated flow direction.

The LK method implementation was provided and used by Barron et al. [4] in their research. The resolution of this field is one vector per pixel so appropriate sub-sampling is used for the numerical comparisons. The sub-sampled LK field corresponding to the previous examples can be seen on Fig. 6.

3. EXPERIMENTS

All mpeg sequences were encoded using an MVCast Mpeg-2 A/V Encoder. Image size is 704x480 with 4:2:0 chroma format and IBBPBBPBBPBB GOP format. A total of 36 video sequences with an average duration of 4 seconds were tested. These were obtained from 9 different surveillance cameras that vary in camera angle, pedestrian density and illumination conditions so results could generalize to a vast range of situations. Table 1 summarizes the error measurements obtained for the most significant compressed domain approximations.

	Total P-frames		*Total I-frames*		*Combined Total*	
16x16	MSE	MAE	MSE	MAE	MSE	MAE
Raw	3,95	1,52				
Raw+Med	0,89	1,35				
Rules	4,41	1,54	2,76	1,38	3,59	1,46
Rules+Med	0,96	1,35	0,93	1,20	0,95	1,28
Rules+Conf	3,02	1,62	1,77	1,33	2,40	1,48
Rules+M+C	0,81	1,44	0,81	1,11	0,81	1,28
8x8	MSE	MAE	MSE	MAE	MSE	MAE
Rules	20,2	0,86	14,5	0,76	17,3	0,81
Rules+Med	5,97	0,58	5,70	0,60	5,84	0,59
Rules+Conf	5,52	0,82	2,92	0,77	4,22	0,80
Rules+M+C	1,59	0,57	1,37	0,62	1,48	0,60

Table 1 – Summary of results obtained

Raw motion vectors (*Raw*) are compared to normalized vectors (*Rules*) using either spatial median filtering (*Med* or *M*) and/or confidence thresholding (*Conf* or *C*). Graph representations of low-resolution (*16x16*) results can be seen in Figs. 7 and 8 while high-resolution (*8x8*) results are shown in Figs. 9 and 10.

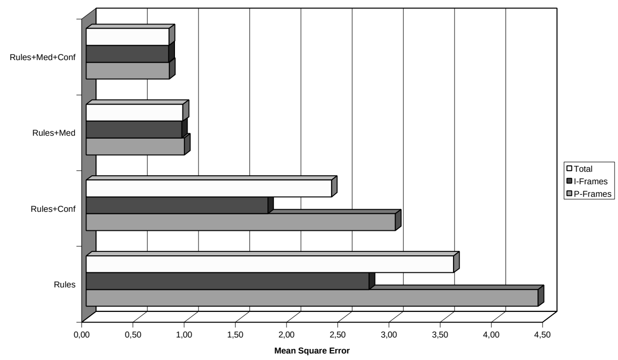

Figure 7 – MSE measurements for low resolution (16x16)

Analysis of these results allows several conclusions. Spatial median filtering reduces MSE measurements significantly both in low and high resolutions. This noise reduction effect is allied with an increased angular accuracy as MAE measurements demonstrate. Confidence maps only improve MSE measurements. They are especially important in high-resolution approximations where they are even more efficient than spatial median filtering. Another interesting result is that approximations errors actually increase if only normalization rules are applied to raw motion vectors. This conclusion is limited to P-frames so it should be carefully considered.

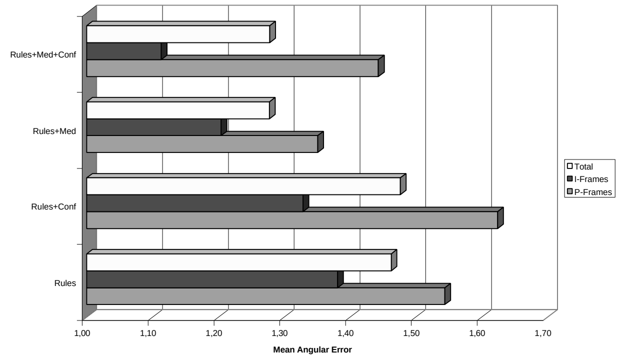

Figure 8 – MAE measurements for low resolution (16x16)

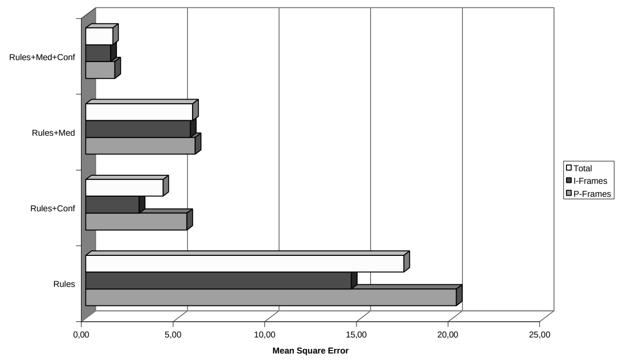

Figure 9 – MSE measurements for high resolution (8x8)

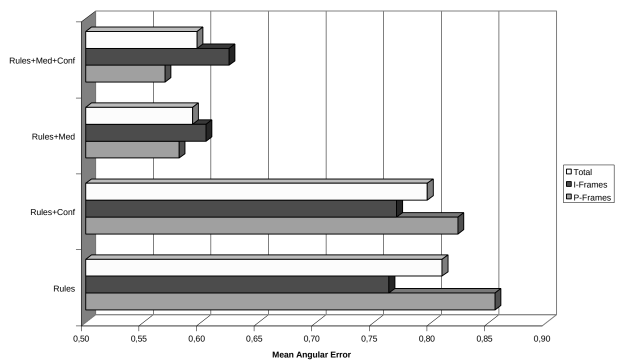

Figure 10 – MAE measurements for high resolution (8x8)

4. DISCUSSION

As a conclusion, *normalization rules* are required to obtain motion fields with one vector per location but they increase noise slightly when compared to *raw motion vectors*. *Spatial median filtering* is an essential step since it drastically reduces noise and improves vector angular accuracy. *Confidence thresholding* also reduces noise by discarding inaccurate vectors, especially if high resolutions are required. Combining both these techniques MSE reductions rates of 78% and 92% are obtained for low and high resolutions.

Future research includes using confidence maps with not only gradient strength data but also direction information, thus possibly allowing an increased angular accuracy of the optical flow approximation obtained.

ACKNOWLEDGMENTS

This research has been supported by Fundação para a Ciência e Tecnologia. The authors would like to thank Sergio Velastin from Kingston University and all the members of the Queen Mary College DSP group for the useful discussions and comments.

REFERENCES

[1] V. Kobla, and D. Doermann, “Compressed Domain Video Indexing Techniques using DCT and Motion Vector Information in MPEG Video”, in *Proc. of SPIE Conf. On Stor. Retr. Image and Video Datab. V, vol. 3022, pp. 200-211, 1997.*

[2] M. Pilu, “On Using Raw MPEG Motion Vectors to Determine Global Camera Motion”, in *SPIE VCIP*, vol. 3309, 1998.

[3] M. Coimbra, M. Davies, and S. Velastin, “Pedestrian Detection using MPEG-2 Motion Vectors”, in *Proc. of WIAMIS 2003*, London, pp. 164-169, 2003.

[4] J.L. Barron, D.J. Fleet, and S.S. Beauchemin, “Performance of Optical Flow Techniques”, in *International Journal of Computer Vision*, vol. 12, pp. 43-77, 1994.

[5] B. Galvin et al. “Recovering Motion Fields: An Evaluation of Eight Optical Flow Algorithms”, in *Proc. of BMVC*, pp. 195-204, 1998.

[6] B. Lucas, and T. Kanade, “An Iterative Image Registration Technique with an Application to Stereo Vision”, in *Proc. of the Imaging Understanding Workshop*, pp. 121-130, 1981.

[7] ISO/IEC JTC1 IS 13818 - 2 (MPEG-2), 1996.

[8] M. Coimbra, and M. Davies, “A New Pedestrian Detection System using MPEG-2 Compressed Domain Information”, in *Proc. IASTED VIIP*, Malaga, Spain, pp. 598-602, 2003.