COMPRESSED DOMAIN GLOBAL MOTION ESTIMATION USING THE HELMHOLTZ TRADEOFF ESTIMATOR

Michael Tok, Alexander Glantz, Marina Georgia Arvanitidou, Andreas Krutz, Thomas Sikora

Communication Systems Group
Technische Universität Berlin
Berlin, Germany

ABSTRACT

Several algorithms for global motion estimation in video sequences using pixel- or block-based approaches have been published. Most known pixel-based methods lack in performance while when using block-based algorithms working on motion vectors, robustness to outliers and accuracy is missing. In this paper we present the fundamentals of a significantly improved, robust block-based method for global motion estimation in compressed domain following the generic Helmholtz principle. To this aim, we use motion vector fields as provided by MPEG data streams. Background PSNR values for four motion compensated test sequences show that our new method delivers results comparable to more complex algorithms.

Index Terms— Global Motion Estimation, Helmholtz Tradeoff Estimator, compressed domain, robust regression

1. INTRODUCTION

Global Motion Estimation (GME) for describing the motion relation between two video frames is a key technique in several video analysis tasks, i.e. background sprite generation, hybrid video coding, etc. Most published approaches are pixel-based as presented in [1], [2] or block-based using motion vector fields from MPEG data streams or other sources as described in [3]. Pixel-based methods are said to deliver more accurate results than motion vector-based approaches. However, heavy computation load is their main drawback. Furthermore they cannot be used for GME on encoded video streams as they need pixel data.

Block-based GME methods on the other hand do not need pixel data as they are employing motion vectors solely. Hence, when working on macroblock structures delivered by MPEG data streams which contain motion vector information, there is no need to decode these video streams completely in order to get a global motion description between two frames [3]. On the other hand, these motion vector fields are noisy, meaning they provide a high percentage of motion vectors not belonging to global motion. Detecting and removing these outliers is the key task for robust regression methods as also described in [4] and [5].

The block-based algorithms introduced in this paper use a generic estimator concept based on the Helmholtz principle as described in [6]. As this estimator is highly robust, GME can be done with high accuracy. Thus, a higher order motion model with precision comparable to results of more complex algorithms can be estimated. One approach described in [7] uses the Helmholtz Tradeoff Estimator (HTE) on lines placed on motion vector positions as a first step for GME. Nevertheless, this two-step method leads to misestimations in presence of larger foreground objects as it takes only few selected motion vectors out of a complete motion vector field.

In this paper we introduce a new method using the HTE in a comparably efficient way with much higher robustness and so more reliable results, where robustness means the ability of rejecting outliers in an environment with only few reliable motion vectors belonging to global motion. Therefore, an optimization step for the HTE using two differently complex motion models is presented.

This paper is organized as follows. First, Section 2 gives a short description of how the Helmholtz Tradeoff Estimator works and how it can be used for GME efficiently. Section 3 outlines an optimization step for the HTE using differently complex motion models to reduce calculation load for GME. For comparison, the first GME method with HTE as shown in [7] is described shortly in Section 4. Experimental evaluation of the described methods and a comparison to a pixel-based approach concerning background PSNR curves and mean values can be found in Section 5. Section 6 summarizes this paper.

2. GME USING HTE DIRECTLY

The Helmholtz Tradeoff Estimator is a robust estimator with the ability of detecting up to 80% of outliers out of a given dataset for an underlying model using subsets, similar to the random sample consensus (RANSAC) [5] or Least Median

The research leading to these results has received funding from the European Community's Seventh Framework Programme [FP7/2007-2011] under grant agreement no. 216444

of Squares implementation described in [4]. The minimal amount of needed subsets m can be calculated by

$$m = \frac{\log(1-P)}{\log(1-(1-\epsilon)^p)}, \tag{1}$$

where P is the desired probability of finding a good estimation in an environment with an outlier percentage of at most ϵ for a model with p parameters. With about $\epsilon = 50\%$ outlier tolerance, which still is high enough for sequences as *Foreman* and $P = 95\%$ assurance for a good result, $m = 766$ subsets of four motion vectors are needed for an eight parameter transformation model as the perspective homography. For every subset out of a given motion vector field taken randomly, a perspective transformation matrix H is calculated by

$$A \cdot \begin{pmatrix} m_0 \\ m_1 \\ \vdots \\ m_6 \\ m_7 \end{pmatrix} = \begin{pmatrix} x_1^{'} \\ y_1^{'} \\ \vdots \\ x_4^{'} \\ y_4^{'} \end{pmatrix}, \tag{2}$$

with

$$A = \begin{pmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1 x_1^{'} & -y_1 x_1^{'} \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1 y_1^{'} & -y_1 y_1^{'} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_4 & y_4 & 1 & 0 & 0 & 0 & -x_4 x_4^{'} & -y_4 x_4^{'} \\ 0 & 0 & 0 & x_4 & y_4 & 1 & -x_4 y_4^{'} & -y_4 y_4^{'} \end{pmatrix} \tag{3}$$

and finally

$$H = \begin{pmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ m_6 & m_7 & 1 \end{pmatrix}. \tag{4}$$

In the next step, for each subset s every motion vector position $(x_i, y_i)^T$ is transformed to $(x_i^{'}, y_i^{'})^T$ by this homography H using

$$\begin{pmatrix} x_i^{'} \cdot h \\ y_i^{'} \cdot h \\ h \end{pmatrix} = H \cdot \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix}. \tag{5}$$

The λ-th percentile $\nu_{s,\lambda}$ of the distances

$$r_{s,i}^2 = (x_i^{'} - \tilde{x}_{s,i})^2 + (y_i^{'} - \tilde{y}_{s,i})^2 \tag{6}$$

from the compensated positions $(\tilde{x}_i, \tilde{y}_i)^T$ to the transformed positions $(x_i^{'}, y_i^{'})^T$ is taken to estimate a standard deviation

$$\sigma_s = 1.4826 \cdot [1 + (5/(n-p))] \cdot \nu_{s,\lambda}. \tag{7}$$

Here, λ is defined by $\lambda = 1 - \epsilon$. The motion vectors are now classified by their residual values with

$$w_{s,i} = \begin{cases} 1, & \text{if } r_i^2 \leq (\frac{5\sigma}{2})^2 \\ 0, & \text{else} \end{cases} \tag{8}$$

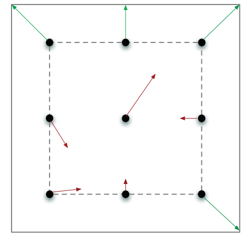


Fig. 1. Direct homography estimation only using inlier motion vectors while rejecting a high amount of outliers

where $w_{s,i} = 1$ classifies a motion vector as inlier. All inliers are used to calculate a final homography H_s by least squares for every subset. With the amount of inliers

$$I_s = \sum_{i=1}^{n} w_{s,i} \tag{9}$$

and their standard deviation

$$\sigma_s^{'} = \sqrt{\sum_{k \in \text{Inliers}} \frac{(r_k - \mu_s)^2}{I_s}} \tag{10}$$

related to H_s with μ_s as mean error of the set, a rating value

$$\Phi_s = \frac{I_s}{\sigma_s^{'}} \tag{11}$$

can be defined for every subset. By taking the homography belonging to the subset with the highest amount of inliers and lowest variance and therefrom having the highest rating Φ_s, the homography H_s representing global motion best is selected. Now, a homography can be estimated reliably even in an environment with a high amount of outliers as Fig. 1 illustrates.

3. SUBSET AMOUNT REDUCTION

As a higher order motion model with more parameters needs more subsets to ensure a good estimation, while lower order motion models deliver worse descriptions of global motion, a combination of two models of different complexity can be used. Thus, for removing outliers with the Helmholtz Tradeoff Estimator in the first step, a simple model can be used to reduce the amount of needed subsets enormously. A model which describes translation, rotation and zoom needs only four parameters and is complex enough to describe the relation between two frames approximately. A homography for such a model can be defined by

$$H = \begin{pmatrix} m_0 & m_1 & m_2 \\ -m_1 & m_0 & m_3 \\ 0 & 0 & 1 \end{pmatrix}. \tag{12}$$

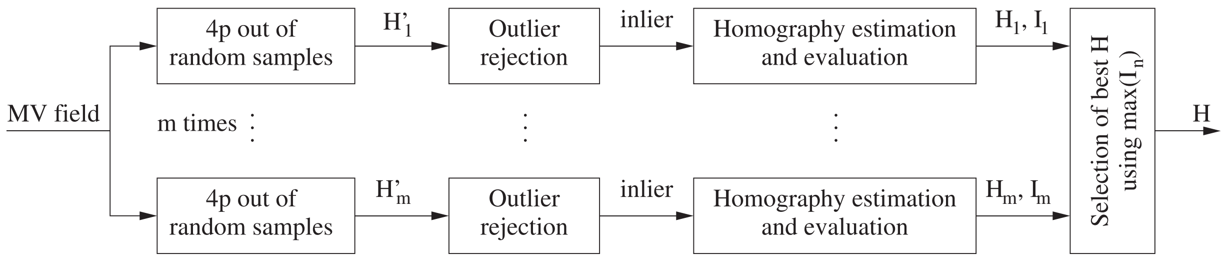


Fig. 2. Global Motion Estimation algorithm using the Helmholtz Tradeoff Estimator and two different motion models

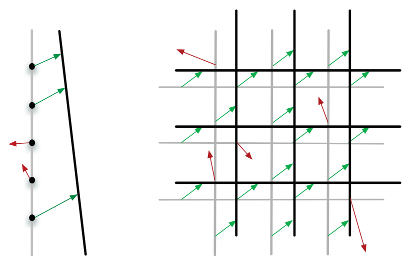


Fig. 3. Line estimation out of potentially representative motion vectors with outlier rejection for line tracking

	Allstars	Biathlon	Foreman	Stefan
uncompensated	30.75	24.20	27.82	17.78
6 line tracking	41.02	32.35	31.80	28.12
8 line tracking	40.86	34.56	30.38	28.83
direct HTE	**42.3**	**39.14**	**37.61**	**30.26**
pixel-based	41.65	37.69	37.32	30.58

Table 1. Mean BPSNR values [dB] for GME

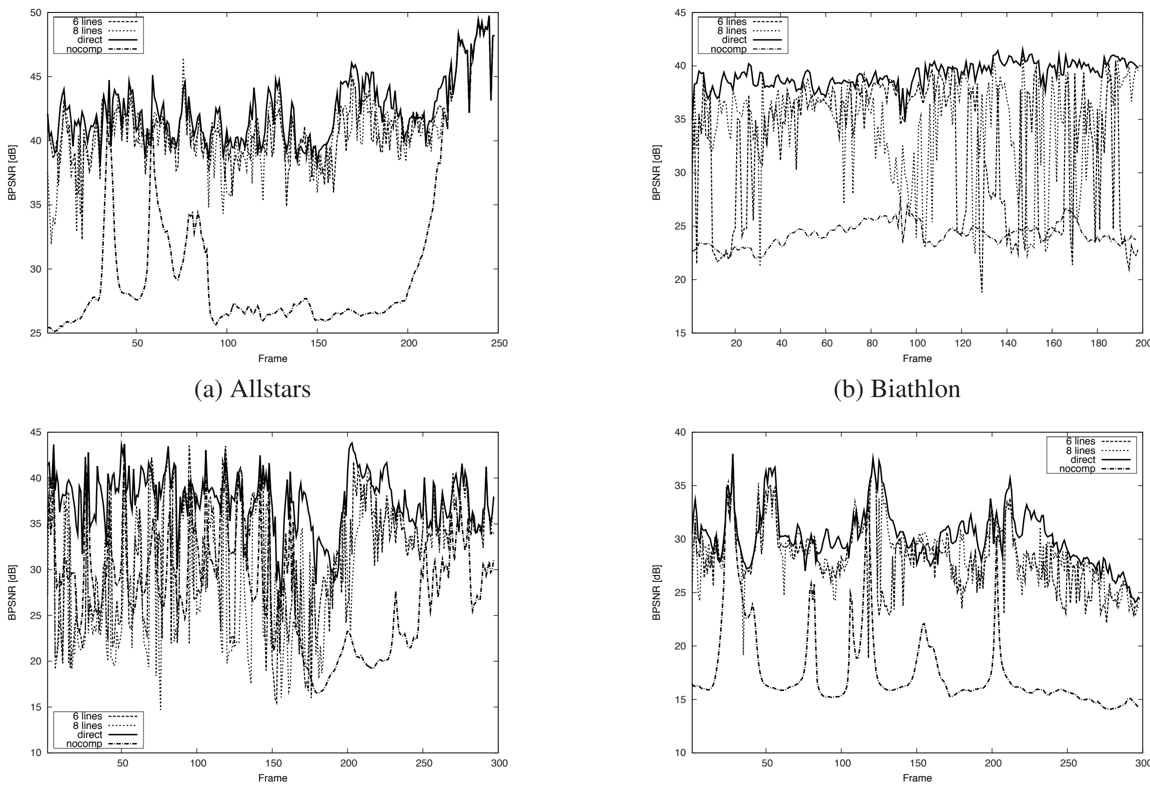
So with even more robust demands as $P = 99.5\%$ ensurance for a good result and $\epsilon = 70\%$ of expected outlier appearence, only $m = 652$ subsets are needed for outlier rejection as reported in Eqn. (1). With the remaining inliers, an eight parameter homography still can be calculated. Thus, higher robustness without lack of accuracy is achievable. Fig. 2 shows the final algorithm.

4. ROBUST GME BASED ON LINE TRACKING

In the first algorithm that used the Helmholtz Tradeoff Estimator for GME [7], the homography parameters were calculated by first estimating lines out of motion vectors and then calculating perspective transformations out of these lines. A line l_i in implicit representation $a_i x + b_i y + c_i = 0$ can also be described perspectively by a vector $l_i = (t_i, u_i, 1)^T$ with $t_i = a_i/c_i$ and $u_i = b_i/c_i$. Considering Eqn. (1), the motion estimation for a line needs much less subsets than the estimation of an affine or perspective tranformation, when desiring a high assurance for a good model in a noisy environment. Thus, applying the HTE on motion vectors for estimating lines reduces computation effort when using high parameter values for P and ϵ. In this way, only motion vectors on selected lines are taken and most of them are not used for estimation. The original algorithm estimates three horizontal and three vertical lines $\tilde{l}_1 \ldots \tilde{l}_6$ as seen in Fig. 3. With the resulting line correspondencies, a homography H can be calculated with $\tilde{l}_i = (H^{-1})^T l_i$. As only four line correspondencies are needed to calculate a perspective transformation, Least Median of Squares, a robust estimation technique described in [4], is applied on the six lines. As Least Median of Squares can reject up to 50% of outliers out of a data set, but four out of six lines already are needed to estimate a homography, we found that the usage of eight lines often is even more robust. Nevertheless, misestimation of only a few lines out of a small dataset consisting only of six or eight lines very likely leads to misestimation of a whole homography.

5. EXPERIMENTAL EVALUATION

Four test sequences were considered for the experimental evaluation.Table 1 shows background PSNR values for the sequences *Allstars*, *Biathlon*, *Foreman* and *Stefan* calculated by estimating the motion between two frames and compensate it with usage of bicubic spline interpolation and then calculating their background error. BPSNR curves for all sequences are shown in Fig. 4. For getting only the background PSNR, foreground segmentation masks were used. To grant better comparability, PSNR values for error frames without motion compensation are also provided. For GME as preprocessing step for motion compensation we used the original algorithm [7], a modified version using eight lines, our new method and a very accurate pixel-based method using gradient descent [1]. As can be seen, our new method performs best among the selected block-based methods in every shown case concerning the accuracy of frame registration. The improved version of the proposed line tracking algorithm works better than or comparable to the original. However, several drops in the curves for the two motion estimation methods with line tracking reveal completely misestimated transformation matrices for both cases.



(a) Allstars (b) Biathlon

(c) Foreman (d) Stefan

Fig. 4. BPSNR values comparing the quality of motion estimation done by the described methods

6. SUMMARY

We proposed a new robust algorithm for global motion estimation working directly on motion vector data in compressed domain and compared it to a similar existing method in an original and a slightly modified version. We also showed that our new method delivers results comparable to pixel-based algorithms but without the need of access to pixel data.

7. REFERENCES

[1] A. Krutz, M. Frater, M. Kunter, and T. Sikora, "Windowed image registration for robust mosaicing of scenes with large background occlusions," in *IEEE Int. Conf. on Image Processing (ICIP06)*, Atlanta, GA, USA, Oct. 2006.

[2] F. Dufaux and J. Konrad, "Efficient, robust, and fast global motion estimation for video coding," *Image Processing, IEEE Transactions on*, vol. 9, no. 3, pp. 497–501, Mar 2000.

[3] M. Pilu, "On using raw MPEG motion vectors to determine global camera motion," *Digital Media Dept., HP Labs. Bristol, Tech. Rep. HPL-97-102*, Aug. 1997.

[4] P. J. Rosseeuw and A. M. Leroy, "Robust regression and outlier detection," New York: Wiley-Interscience, 1987.

[5] M. A. Fischler and R. C. Bolles, *Random sample consensus: A paradigm for model fitting with applications to image analy- sis and automated cartography*, Communications of the ACM, vol. 24, pp. 381–395, 1981.

[6] R. Felip, X. Binefa, and J. Diaz-Caro, "A new parameter estimator based on the helmholtz principle," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on Computer Vision (ICCV)*, Sept. 2005, vol. 2, pp. II–306–9.

[7] R. Felip, L. Barcelo, X. Binefa, and J. Kender, "Robust dominant motion estimation using MPEG information in sport sequences," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 18, no. 1, pp. 12–22, Jan. 2008.