Digital Signal Processing 46 (2015) 1–9

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Motion vector outlier removal using dissimilarity measure

Burak Yıldırım [a], Hakkı Alparslan Ilgın [b,*]

[a] Undersecretariat for Defense Industries, İnönü Bulvarı, Ankara, 06100, Turkey
[b] Ankara University, Electrical and Electronics Eng. Department, Ankara 06830, Turkey

ARTICLE INFO

Article history:
Available online 10 August 2015

Keywords:
Global motion estimation
Motion vector
Similarity measure
Dissimilarity measure
Outlier removal

ABSTRACT

Global motion estimation, being one of the most important tools in video processing field with many applications, is mainly carried out in pixel or compressed domain. Since those based on the pixels have drawbacks such as high computational complexity, most researches are oriented to the compressed domain in which motion vectors are utilized. On the other hand, there are many unwanted existing outliers in motion vector based global motion estimation because of noise or foreground effects. In this paper, proposed motion vector dissimilarity measure is used to remove the outliers to provide fast and accurate motion vector based global motion estimation. Performance of the dissimilarity measure is further improved by using different neighborhood orientations. Also phase correlation of motion vectors are effectively utilized. Therefore small noisy motion vectors are easily detected and different orientations contribute to both performance and low latency. Experiments using the proposed method achieve more accurate results with less computational complexity compared to the state of the art methods.

1. Introduction

Global Motion Estimation (GME) is a common process of camera motion modeling by selecting and estimating appropriate transform parameters in video sequences [1,2]. GME has been widely used in many applications such as video retrieval and indexing, image registration, background modeling, moving object segmentation, scene analysis, object-oriented coding, and MPEG-7 video descriptors. Video stabilization by removing undesired motions with saliency constraints is implemented in [3]. In this particular work, stabilization is based on computing camera paths directed by a variety of automatically derived constraints in which RANSAC is used with a fast grid-based approach. Another global motion based method is implemented for video stabilization using homography consistency in which smooth trajectories and consistent inter-frame transition are obtained [4].

Conventional pixel-based GME provides a good performance, but have high computational cost especially for real time applications [5]. Therefore, researches in this field are focused on the compressed video sequences in terms of bitstream. Extraction of motion vectors (MVs) from bitstream allows a fast GME when compared to the pixel domain. However, MVs in the compressed domain are often imperfect and incongruous with real camera motion [6]. But they are still suitable for removal of outliers with satisfying results. In [7], pixel and MV based GME are evaluated for camera motion characterization.

In MV based GME, coarsely sampled MV fields from the compressed video are utilized as input data for GME [6]. In recent studies, different methods are used for both outlier removal and GME. In [1], Newton–Raphson Gradient Descent (GD) method is proposed with an iterative approach to estimate the global motion from a coarsely sampled MV field. In [5], tensor voting is used for outlier removal whereas it uses GD for GME. MV based outlier removal is commonly accomplished with MVs by calculating spatial magnitude and phase correlations with their neighbors. Nguyen and Lee have proposed tensor voting for outlier removal before GME [8]. According to their approach (TV_GD), MVs are first encoded by second order stick tensors. A 2D stick voting process is then used for each tensor to its predefined neighborhood using a stick kernel. In voting process, new tensors are encoded and simply summed. Finally, by decomposition process of final tensors, new eigenvectors are obtained. Input MV field and these eigenvectors are compared by a similarity, which uses a phase difference to remove outliers. In [6] (CAS_GD), a cascade composed of three-filters for outlier removal is used. Input MV is compared with its different oriented neighbors by both magnitude and phase difference to detect outliers in each filter. In [9] (FLT_GD), a magnitude difference between a MV and its eight neighbors are used. In [10] (LSS_ME), least square solution with M-estimator is used in order to reduce influence of outlier MVs and estimate global motion. In [11] (LMedS), after prefiltering of MVs, least median of squares

* Corresponding author.
E-mail addresses: byildirim@ssm.gov.tr (B. Yıldırım), ilgin@eng.ankara.edu.tr (H.A. Ilgın).

http://dx.doi.org/10.1016/j.dsp.2015.08.002

Table 1
Explanation of notations.

Notation	Explanation
$\boldsymbol{a}$	Global motion parameters
$\boldsymbol{\Delta a}$	Change of global motion parameters
(x_i, y_i)	Coordinate of the ith block
(MVx_i, MVy_i)	Motion vector of the ith block
$(\Delta x_i, \Delta y_i)$	Displacement vector for the ith block
E	Total estimation error
s_k	Step size
g_k, g_p	Gradient vector
H_k, H_{pr}	Hessian matrix
$MV(i, j)$	Input motion vector
$MV(k, l)$	Neighbor motion vector
R	Neighbor coordinates of the interested block
$\|\boldsymbol{u}\|$	Norm of vector u
φ	Phase difference between two MVs
$d(i, j\|n)$	nth dissimilarity value for block (i, j)
$dS(i, j)$	Total dissimilarity value for block (i, j)
V_r	Voting range
σ	Standard deviation
c	Function of scale

estimator is used for both detecting outliers and estimation of global motion. This method has high computational cost due to the decoding of DCT coefficients and randomly chosen parameters of camera motion. In [12], Random Sample Consensus with least square regression (RAN_LS) is used for GME [8,12].

In this paper MV dissimilarity measure is proposed to remove outliers to achieve fast and accurate MV-based GME. GME with different motion models and GD algorithms are explained in the following sections. Proposed MV outlier removal approach is detailed in Section 4. Experimental comparison of the proposed approach with other state of the art methods is given in Section 5. The last section concludes the paper with final remarks. Variables used in this study and their explanations are summarized in Table 1.

2. Camera motion modeling

Global motion models, such as translational, geometric, affine and perspective, are widely used in modeling of camera motion [13]. Perspective model has eight parameters being the most general one among others as described in [5]. It is taken as a basis in [6,8] to generate motion vector field and perform GME. Affine model is also popular and has lower computational cost, which is appropriate for real time applications. In our approach, we use the perspective model similar to the one in [6,8]. For the perspective model, parameters can be defined as $a = [a_0, a_1, \ldots, a_7]$. These parameters are then estimated to obtain global motion. Perspective model transformation then can be defined as follows:

$$\begin{aligned} x' &= f_x(x, y|a) = \left(\frac{a_0 x + a_1 y + a_2}{a_6 x + a_7 y + 1}\right) \\ y' &= f_y(x, y|a) = \left(\frac{a_3 x + a_4 y + a_5}{a_6 x + a_7 y + 1}\right) \end{aligned} \tag{1}$$

where (x, y) and (x', y') are the coordinates in the current and reference frames, respectively. GME based on MVs is constitutively performed with an iterative process: Input motion vector field is used to estimate global motion parameters. First of all, these parameters are initialized by assigned values using input MVs. Iterative process is then implemented to estimate the parameters through minimizing the estimation errors between the input and the estimated motion vector field. If the specified condition about the change of the parameters is satisfied, iteration stops and the estimated parameters are regarded as GM parameters. In MV based GME, input frame is divided into blocks. Each block is represented with a motion vector. All MVs in the frame constitute input motion vector field. Considering i as the block index, (x_i, y_i) as the center of the ith block, and (MVx_i, MVy_i) as motion vector of the ith block; the displacement vector for motion model $\boldsymbol{a}$ is given as:

$$\begin{aligned} (\Delta x_i, \Delta y_i) &= \left(x'_i - x_i, y'_i - y_i\right) \\ &= \left(f_x(x, y|a) - x_i, f_y(x, y|a) - y_i\right) \end{aligned} \tag{2}$$

We need to minimize the estimation error between the input and the estimated MV field. Total error E for L motion vectors is calculated as follows:

$$E = \sum_{i=1}^{L} w_i\left(\left(MVx_i - \left(x'_i - x_i\right)\right)^2 + \left(MVy_i - \left(y'_i - y_i\right)\right)^2\right) \tag{3}$$

where binary weight w_i is 0 or 1, when MV is outlier or inlier, respectively. Before GME, outlier MVs, which are opposed to global motion, should be eliminated. Outlier MVs do not fit real camera motion and, the better they are detected and discarded, the better GME is ensured with less iteration, which means faster convergence to reach final estimation parameters. MVs obtained by Block Matching Estimation (BME) algorithm may appear as outliers as follows:

- During block-matching process, if Sum of Absolute Differences (SAD) for a block is high, MV, which belongs to this block, is most probably outlier.
- If MV differs from its neighbors, it is more likely a foreground MV.
- The blocks located at the border of a frame is usually emerged all of a sudden and do not resemble any of its neighbors.

GME performance can be improved with a robust outlier removal algorithm which is the key point of our study.

3. Gradient descent (GD) algorithm

GD algorithm is a well-known optimization method and frequently used in parameter estimation. It can be implemented to one- or multi-dimensional unconstrained problems [14]. The basic principle of GD is to find local minimum of a function. If a function (objective) is in parabola form, its local minimum can be easily calculated with first or second order differentiation. If differential information is not known or the function is not a parabola, GD can be easily exploited. First, an initial guess point is determined, and then this point is decreased/increased with a specified direction and a step size by iteration using objective function. If difference between decreased/increased (updated) and previous value is below than a predefined threshold, iteration stops and updated value is regarded as local minimum point. In our study, the objective function is error E given in Eq. (3). In gradient based optimization, The Newton–Raphson Method [15] is very effective for estimating transform parameters. The biggest benefit of this method is to enable usage of second-order Taylor series expansion of the function about the current design point, i.e. a quadratic model [16]:

$$f(x_k + s_k) \approx f_k + g_k^T s_k + \frac{1}{2} s_k^T H_k s_k \tag{4}$$

where s_k is step (further replaced with transform parameters) to minimum $(x_k + s_k)$. If we differentiate Eq. (4) with respect to s_k, the left side will be zero. The right side of the equation is in quadratic form and consists of gradient vector g_k and Hessian matrix H_k. After the differentiation, we obtain the step (parameter), which minimizes the quadratic:

$$s_k = -H_k^{-1} g_k \tag{5}$$

where s_k is now the solution of the algorithm. In our study, we use GD as in [1] because of the advantages explained to estimate