We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

7,000
Open access books available

186,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Global Motion Estimation and Its Applications

Xueming Qian

School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China

1. Introduction

In this chapter, global motion estimation and its applications are given. Firstly we give the definitions of global motion and global motion estimation. Secondly, the parametric representations of global motion models are provided. Thirdly, global estimation approaches including pixel domain based global motion estimation, hierarchical global motion estimation, partial pixel set based global motion estimation, and compressed domain based global motion estimation are reviewed. Finally, four global motion based applications in video compression, sport video shot classification, video error concealment, and video text occluded region recovery are given.

Motion information is very important for video content analysis. In surveillance video, usually the camera is stationary, and the motions of the video frame are often caused by local motion objects. Thus detecting motions in the video sequences can be utilized in abnormal events detection. In sports video, the heavy motions are also related to highlights. Motion estimation and compensation is the core of video coding. Coding the residual component after motion compensated can save bit-rates significantly. In video sequences, the motion pattern can be classified into two types: local motion and global motion. The global motion is related to camera motion. Integrated with local motion, global motion is widely utilized in video object segmentation, video coding and error concealment. The rest of this chapter is organized as follows: the definition of global motion is given in Section 2. The global motion models are given in Section 3. Global motion estimation approaches are given in Section 4. Four applications based on global motion and local motion (GM/LM) information are introduced in Section 5. The applications are GM/LM based video coding, global view refinement for soccer video, GM/LM based error concealment and GM/LM based text occluded region recovery. And finally conclusions are drawn in Section 6.

2. Definition of global motion

Global motions in a video sequence are caused by camera motion, which can be modeled by parametric transforms [4]. The process of estimating the transform parameters is called global motion estimation.

From the definition, it is clear that global motion is closely related to camera motion. The camera is operated by camera man. Thus the global motion pattern can reveal video

shooting style which has some relationship with video contents [18]. The global motion information is especially useful in sport video content analysis [13]-[18].

From the definition, we find that the global motions have certain consistence for the whole frame as shown in Fig.1. The global motion in Fig.1 (a) is a zoom out and that in Fig.1 (b) is a translation respectively. From Fig.1 (a), we find that the motion direction is from outer to inner regions, which means that the coordinates of a current frame t can be generated in the inner regions of the reference frame v ($t > v$). In Fig.1, the motion vectors in the motion field correspond to the global motion vectors at the coordinates.

Global motion vector is the motion vector calculated from the estimated global motion parameters. Global motion vector $(GMVx_t, GMVy_t)$ for the current pixel with its coordinates (x_t, y_t) is determined as

$$\begin{cases} GMVx_t = x'_t - x_t \\ GMVy_t = y'_t - y_t \end{cases} \tag{1}$$

where (x'_t, y'_t) are the warped coordinates in the reference frame by the global motion parameters from the coordinate (x_t, y_t).

(a)Zoom-out   (b) Translation
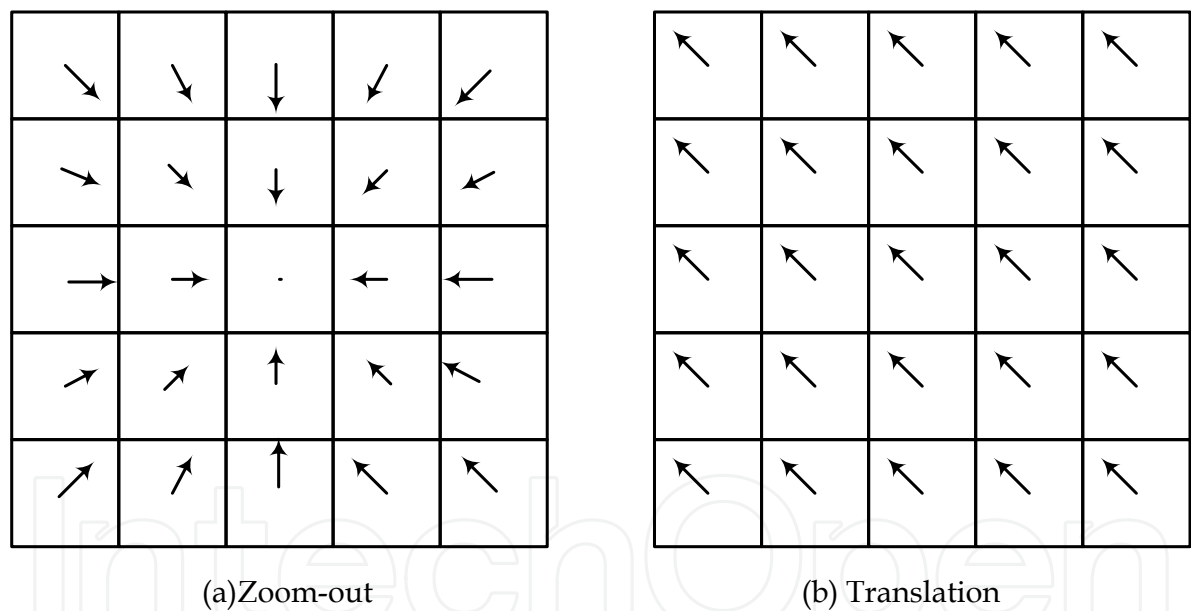
Fig. 1. Global motion fields. (a) Zoom-out and (b) Translation.

3. Global motion models

Global motion can be represented by global motion models with several parameters. The simplest global motion model is translation with only two parameters. The complex global motion model is quadric model with 12 parameters. Generally, higher order models have more parameters to be estimated, which can represent more complex motions. The lower order models are special cases of the higher ones. The widely used global motion model is perspective model with 8 parameters, which is expressed as follows

$$\begin{cases} x' = \dfrac{m_0 x + m_1 y + m_2}{m_6 x + m_7 y + 1} \\ y' = \dfrac{m_3 x + m_4 y + m_5}{m_6 x + m_7 y + 1} \end{cases} \tag{2}$$

where (x, y) and (x', y') are the coordinates in the current and the reference image respectively, with the set of parameters $\mathbf{m} = [m_0, \cdots, m_7]$ denoting the global motion parameters to be estimated. If $m_6 = m_7 = 0$, then it is an affine model with 6 parameters. Then Eq.(2) can be simplified as follows

$$\begin{cases} x' = m_0 x + m_1 y + m_2 \\ y' = m_3 x + m_4 y + m_5 \end{cases} \tag{3}$$

When $m_0 = m_4 = 1$ and $m_1 = m_3 = m_6 = m_7 = 0$, then the perspective model is actually simplified into a translation model as follows

$$\begin{cases} x' = x + m_2 \\ y' = y + m_5 \end{cases} \tag{4}$$

4. Global Motion Estimation (GME) approaches

Intuitively, global motion estimation can be carried out in pixel domain. In the pixel domain based approaches, all the pixels are involved in the estimation of global motion parameters. There are two shortcomings in pixel domain based approach: 1) it is very computational intensive; 2) it is often sensitive to noises (local object motions).

In order to improve the convergence and speed up the calculation, coarse to fine searching approach is often adopted. Moreover, the subset of pixels having the largest gradient magnitude is adopted to estimate the global motion parameters [6]. Sub-point based global motion estimation approaches are very effective in reducing computational costs. To guarantee the accuracy of global motion estimation, how to determine the optimal sub-sets are the key steps. Except the pixel domain based global motion estimation, compressed domain based global motion estimation approaches are also very popular.

Robust global motion estimation usually carries out by identifying the pixels (blocks or regions) that undergo local motions. Fig.2 shows the global motion and local motions. If the local motion blocks can be determined as outliers, then the global motion performance can be improved significantly.

4.1 Pixel domain based GME

In GME involving two image frames I_k and I_v (with $k<v$), one seeks to minimize the following sum of squared differences between I_v and its predicted image $I_k(x(i, j), y(i, j))$ which is obtained after transforming all the pixels in I_k.

$$E = \sum_i \sum_j e(i, j)^2 \tag{5}$$

where $e(i, j)$ denotes the error of predicting a pixel located at (i, j) of frame I_v, by using a pixel at location $[x(i, j), y(i, j)]$ of previous frame I_k.

$$e(i,j) = I_v(i,j) - I_k(x(i,j), y(i,j))\qquad(6)$$

The transform mapping functions $x(i, j)$ and $y(i, j)$ (with respect to global motion parameters $\mathbf{m}$) should be so chosen that E in Eq.(5) is minimized. The well-known Levenberg-Marquard algorithm (LMA) or lest square approach, can be utilized to find the optimal global motion parameters $\mathbf{m}$ iteratively by minimizing the energy function in Eq.(5) as follows

$$\mathbf{m}^{(n+1)} = \mathbf{m}^{(n)} + \Delta\mathbf{m}^{(n)}\qquad(7)$$

where $\mathbf{m}^{(n)}$ and $\Delta\mathbf{m}^{(n)}$ are the global motion parameters and updating vector at iteration n [8].

All the pixels are involved in the global motion parameters optimization in the traditional LMA algorithm [9]. This is very computational intensive. It is impractical for real-time applications. Moreover, the local motions in video frame may also bias the global motion parameters' estimation precision. Thus improvements are carried out by utilizing hierarchical global motion estimation, partial pixel set and compressed domain based approaches.

4.2 Hierarchical global motion estimation

In MPEG-4, GME is performed by a hierarchical approach to reduce computational costs [1]. It is an improvement of pixel domain based approach which consists of following three steps. Firstly, spatial pyramid frames are constructed. Secondly, global motion parameters with the coarsest global motion model are estimated at the top layer of the pyramid images. Then, the estimated global motion parameters at the coarsest level are projected to its next high resolution level to get the refined global motion parameters. Finally, the refined global motion parameters are iteratively updated using a least-square based approach and the process continues until convergence [1]. Fig. 2 shows the illustration of hierarchical global motion estimation approach. The original image and its motion field, the second layer and third layer pyramid images and their motion fields are shown in Fig.2 (a), (b) and (c) respectively. In Fig.2 local motion region (LMR) and global motion region (GMR) of each layer are labeled out respectively.

Global motion parameters at each layer are estimated by minimizing the sum of weighted squared errors over all corresponding pairs of pixels (x_i, y_i) and (x_i', y_i') within the current image f and the reference image R as follow.

$$E = \sum_i \sum_j w(i,j) e(i,j)^2\qquad(8)$$

$$E = \sum_i \sum_j w(i,j)[R(x(i,j), y(i,j)) - f(i,j)]^2\qquad(9)$$

where $w(i, j)$ is the corresponding weight of the pixel at coordinate (i,j) with $w(i,j) \in \{0,1\}$. You know, local object motion may create outliers and therefore bias the estimation performance of the global motion parameters. To reduce the influence of such outliers, a

robust histogram based technique is adopted to reject the pixel points with large matching errors by setting their weights to be "0".

The hierarchical global motion estimation approach has following advantages: 1) estimating the coarse global motion parameters on the top layer of pyramid is effective for noise filtering; 2) computational cost of coarse global motion estimation is very low at the top layer of pyramid. This is due to the fact that only small resolution images are involved in GME and the global motion model is low order which is easy to get convergence; 3) adaptive model determination with respect to the precisions of global motion parameters, which is also helpful for reducing computational cost. In the enhanced layer, it is only need to updating global motion parameters on the basis of the parameters estimated in its previous layers. The advantages of hierarchical global motion over traditional pixel domain based global motion estimation approach can be shown by the illustrations in Fig.2 respectively.
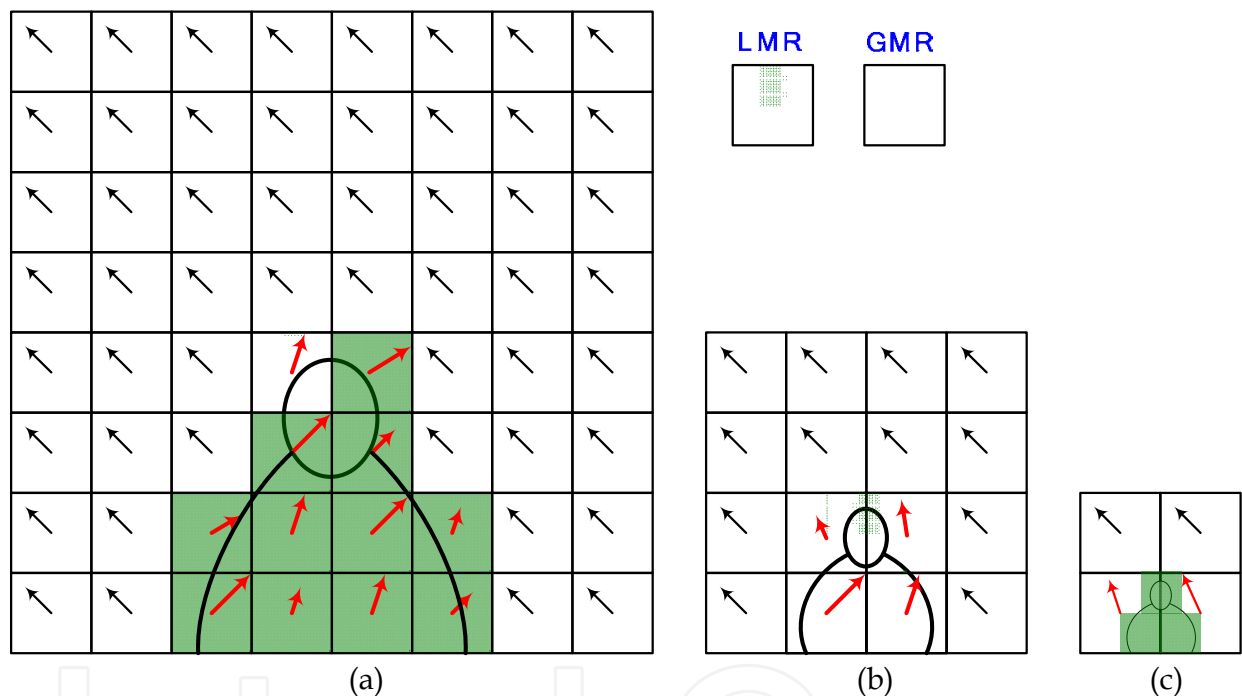

Fig. 2. Illustration of hierarchical global motion estimation approach. (a) Original image and its motion field, (b) and (c) correspond to the second layer and third layer pyramid images and their motion fields.

4.3 Partial pixel points based GME

Just as its name implies, partial pixel points based GME approaches only use sub-set of the whole pixels for estimating global motion parameters. In [6], the subset utilized for GME is selected based on gradient magnitudes information. The top 10% pixels with the largest gradient magnitudes are selected and severed as reliable points for GME. This method divides the whole image into 100 sub-regions and selects the top 10% pixels as feature points which can avoid numerical instability. This subset selection approach reduce the computational cost by reduce the number of pixels at the cost of calculating the gradient image and ranking the gradient of the whole pixels. To further reduce the computational

cost, a random subset selection method was proposed in [4] for GME in fast image-based tracking. Pixel selection can also follow certain fixed subsampling pattern. Alzoubi and Pan apply the subsampling method that combines random and fixed subsampling patterns to global motion estimation [9]. The corresponding combined subsampling patterns can provide significantly improved tradeoffs between motion estimation accuracy and complexity than those achievable by using either fixed or random patterns alone. Wang et al., [7] proposed a fast progressive model refinement algorithm to select the appropriate motion model to describe different camera motions. Based on the correlation of motion model and model parameters between neighbor frames, an intermediate-level model prediction method is utilized.

4.4 Compressed domain based GME

In video coding standards, the motion estimation algorithms calculate the motions between successive video frames and predict the current frame from previously transmitted frames using the motion information. Hence, the motion vectors have some relationship with the global motion [10]-[12]. A global motion estimation method is proposed based on randomly selected MV groups from motion vector field with adaptive parametric model determination [5]. A non-iterative GME approach is proposed by Su *et al.* by solving a set of exactly-determined matrix equations corresponding to a set of motion vector groups [4]. Each MV group consists of four MVs selected from the MV field by a fixed spatial pattern. The global motion parameters for each of the MV group are obtained by solving the exactly-determined matrix equation using singular value decomposition (SVD) based pseudo-inverse technique. The final global motion parameters are obtained by a weighted histogram-based method. Moreover, a least-square based GME method by coarsely sampled MVs from the input motion vector field is proposed for compressed video sequences [5]. The global motion parameters are optimized by minimizing the fitting error between the input motion vectors and the wrapped ones from estimated global motion parameters. In order to estimate global motions robustly, motion vectors in local motion region, homogeneous region with zero or near-zero amplitude and regions with larger matching errors are rejected.

The objective function of compressed domain based GME approaches is to minimize the weighted mean matching error (MME) of the input motion vectors and the generated ones by virtue of the estimated global motion parameters, which is expressed as follows

$$MME = \sum_{i=1}^{MVNum} w_i(ex_i^2 + ey_i^2) \bigg/ \sum_{i=1}^{MVNum} w_i \tag{10}$$

$$\begin{cases} ex_i = MVx_i - x_i' + x_i \\ ey_i = MVy_i - y_i' + y_i \end{cases} \tag{11}$$

where (MVx_i, MVy_i) denotes the input MV of the *i*-th macro-block (MB) at the spatial coordinates (x_i, y_i), (ex_i, ey_i) denote the errors vector between the decoded MV and the MV generated by the estimated global motion parameters. And let (x_i', y_i') denote the warped coordinates for (x_i, y_i) with respect to the global motion parameters **m**. *MVNum* denotes the

number of MVs and $w_i(1 \le i \le MVNum)$ is the weighting factor for the *i*-th MB, with $w_i \in \{0,1\}$. How to reject the outlier motion vectors is also very important to improve global estimation performances [10]. Intuitively, w_i can be set to be "0" if one of the following three conditions is satisfied: 1) this MB is located in a smooth region (which can be indicated by the standard deviation of the luminance component), 2) the matching error of this MB is large enough (which can be measured by the DC coefficient of the residual component), 3) this MB is intra-coded. Global motion estimation is carried out using the MBs with their weights set to be "1".

5. Applications of global motion estimation

In this Section, four global motion based applications are illustrated. They are 1) the GMC (global motion compensation) and LMC (local motion compensation) based video coding in MPEG-4 advanced simple profile (ASP), 2) GM and LM based mid-level semantic classification for sport video, 3) GM/LM based video error concealment, and 4) GM/LM based text occluded region recovery.

5.1 GMC and LMC based video coding

The aim of this part is to illustrate how video compression performances can be improved by utilizing adaptive GMC/LMC mode determination. GMC/LMC based motion compensation mode selection approach in MPEG-4 is given [1], [2]. Global motion estimation and compensation is used in MPEG-4 advanced simple profiles (ASP) to remove the residual information of global motion. Global motion compensation (GMC) is a new coding technology for video compression in MPEG-4 standard. By extracting camera motion, MPEG-4 coder can remove the global motion redundancy from the video. In MPEG-4 ASP, each macro block (MB) can be selected to be coded use GMC or local motion compensation (LMC) adaptively during mode determination. Intuitively, some types of motion, e.g., panning, zooming or rotation; could be described using one set of motion parameters for the entire VOP (video object plane). For example, each MB could potentially have the exact same MV for the panning. GMC allows the encoder to pass one set of motion parameters in the VOP header to describe the motion of all MBs. Additionally MPEG-4 allows each MB to specify its own MV to be used in place of the global MV.

In MPEG-4 Advanced simple profile, the main target of Global Motion Compensation (GMC) is to encode the global motion in a VOP (video object plane) using a small number of parameters. Each MB can be predicted either from the previous VOP by global motion compensation (GMC) using warping parameters or from the previous VOP by local motion compensation (LMC) using local motion vectors as in the classical scheme. The selection is made based on which predictor leads to the lower prediction error. In this Section we only expressed the GMC/LMC mode selection approach. More detail expression for the INTER4V/INTER/field prediction, GMC/LMC, and INTRA/INTER can be found in the Section 18.8.2 GMC prediction and MB type selection [2]. The pseudo-code of GMC/LMC mode decision in MPEG-4 AS is as follows:

if (SAD_{GMC} - P < SAD_{LMC}) then GMC
else LMC

where SAD_{GMC} is defined as the sum of the absolute difference in the luminance block when using GMC prediction, and SAD_{LMC} and P are defined as:
if the previous criterion was INTER4V

$$SAD_{\text{LMC}} = SAD_8$$
$$P = N_B * \text{Qp} / 16$$

if the previous criterion was INTER

$$SAD_{\text{LMC}} = SAD_{16}$$
$$P = N_B * \text{Qp} / 64$$

if the previous criterion was INTER and the motion vector was (0,0)

$$SAD_{\text{LMC}} = SAD_{16} + (N_B / 2 + 1)$$
$$P = NB / 2 + 1$$

if the previous criterion was field prediction

$$SAD_{\text{LMC}} = SAD_{16*8}$$
$$P = N_B * \text{Qp} / 32$$

where SAD_8 (sum of absolute difference for four 8x8 luminance blocks when the INTER4V mode is selected), SAD_{16} (sum of absolute difference for a 16x16 luminance block when the INTER mode is selected) and SAD_{16*8} (sum of absolute difference for two 16x8 interlaced luminance blocks when the field prediction mode is selected) are computed with half pixel motion vectors. N_B indicates the number of pixels inside the VOP. Qp is the quantization parameter.

5.2 Global motion based shot classification for sport video

In [13], Xu et al. classified soccer video shots into the views of global, zoom-in and close-up. From the view sequences, each soccer video clip is classified into either a play or a break. In [14], Duan et al. classified video shots into eight categories by fusing the global motion pattern, color, texture, shape, and shot length information in a supervised learning framework. Ekin and Tekalp utilized shot category and shot duration information to carry out play-break detection according to the dominant related rules and soccer video production knowledge [16]. Similarly, Li et al. classified video shots into event and non-event by identifying the canonical scenes and the camera breaks [15]. Tan et al. also segmented a basketball sequence into wide angle, close-up, fast-break and possible shoot-at-the-basket using motion information [17].

In soccer video, the global views give audiences an overall view of the sport, while the close up and medium views, being complementary to global views, show certain details of the game. Typically the camera men operate cameras, by fast track, or zoom in to provide audiences with clearer views of the games. Based on the view type, camera motion patterns and domain related knowledge, high level semantics can be inferred. The classified shot category information is helpful for highlight events discrimination. In [18], global views of soccer video are further refined into the following three types: stationary, zoom and track in terms of camera motion information using a set of empirical rules with respect to domain and production knowledge. The key-frames of a shot with stationary by means of average motion

intensity and average motion intensities of global motion. The local motion information is represented by average motion intensity (AMV) which is expressed as follows

$$AMV = \frac{1}{M}\sum_{j=1}^{M}\sqrt{MVx_j^2 + MVy_j^2} \tag{12}$$

where (MVx_j, MVy_j) is the motion vector (MV) of the block with its coordinates (x_j, y_j) and j is the block index. The average global motion intensity ($AGMV$) is calculated as follows:

$$AGMV = \frac{1}{M}\sum_{j=1}^{M}\sqrt{GMVx_j^2 + GMVy_j^2} \tag{13}$$

where ($GMVx_j$, $GMVy_j$) is the global motion vector of the block at (x_j, y_j), M is the total block number. The global motion vector $(GMVx_t, GMVy_t)$ at the coordinates (x_t, y_t) is determined as follows

$$\begin{cases} GMVx_t = x_t' - x_t \\ GMVy_t = y_t' - y_t \end{cases} \tag{14}$$

where (x_t', y_t') are the warped coordinates in the reference frame by the global motion parameters from the coordinate (x_t, y_t).

The GM/LM based global view refinement is carried out by the following empirical rules [18]: If the motion energy of a frame satisfies AMV<0.5, then it is stationary otherwise non-stationary. The non-stationary shot is further classified into zoom and track. A frame is a zoom-in if $m_0=m_5$ >1, a zoom-out if $m_0=m_5$ <1, otherwise a track ($m_0=m_5$ =1). The track is a slow-track if the average global motion intensity $AGMV$ satisfies $AGMV \leq 2$, otherwise a fast-track.

5.3 Global motion and local motion (GM/LM) based application in error concealment

The aim of this sub-chapter is to show how combine global and local motion to improve visual video qualities of corrupted video sequences.

5.3.1 Related work on error concealment (EC)

Temporal recovery (TR) is often utilized to replace the erroneous macro-blocks (EMBs) by their spatially corresponding MBs in the reference frames. TR is efficient for the stationary video sequences. Temporal average (TA) uses the average or medium MV of the correctly received MBs in its neighbors to substitute the losing MVs for the corrupted MBs [19]. Boundary pixels of the top and bottom-, or (and) left and right-adjacent MBs as the references [20], [21]. A recursive block matching (RBM) technique is utilized to recover the error MBs [20]. The correctly received MBs in its neighbors are utilized. Recovery results of the corrupted MBs are improved step by step using the full searching technique within a given searching range. However, this approach is not effective when the reference blocks located in texture-alike or smooth region. There are more than one best matches for the two 8×16 blocks in the smooth regions at reference frames.

A global motion based error concealment method is proposed by Su *et al.*[3,4]. In [3] MVs generated by global motion parameters are utilized to recover the EMBs under the assumption that they are all located in global motion regions. When the EMBs are in LM or GM/LM overlapped regions, usually the MVs generated by global motion parameters are incorrect to recover the lost MVs.

5.3.2 GM/LM based error concealment

Obviously it is more effective to recover MVs of the EMBs in the global motion regions by the global MVs and the EMBs in local motion regions by the local motion compensation. And for the corrupted MBs located in the GM/LM overlapped regions, more accurate boundaries need to be searched using the advanced boundary matching criteria [19]-[21]. We give the detailed steps for the GM/LM based error concealment approach [22]. The detail diagram of the proposed GM/LM based EC method is shown in Fig.3.

GM/LM based EC method consists of the following four steps: 1) Carry out global motion estimation for the corrupted frames using the MVs of the correctly received MBs (CMBs). 2) Classify the CMBs into global motion MBs (GMBs) or local motion MBs (LMBs) types. 3) Determine the type of the erroneous MBs (EMBs) and Step 4. Carry out recovery by using the GM/LM based approach.

Based on the estimated global motion parameters, a CMB is classified into two types: GMB and LMB adaptively with respect to the matching error of the reconstructed MB (from the video streams) and the global motion warped MB. If the matching error is large enough then it is a LMB, otherwise a GMB. Actually this step does not influence the GM/LM based error concealment performances very much [22].
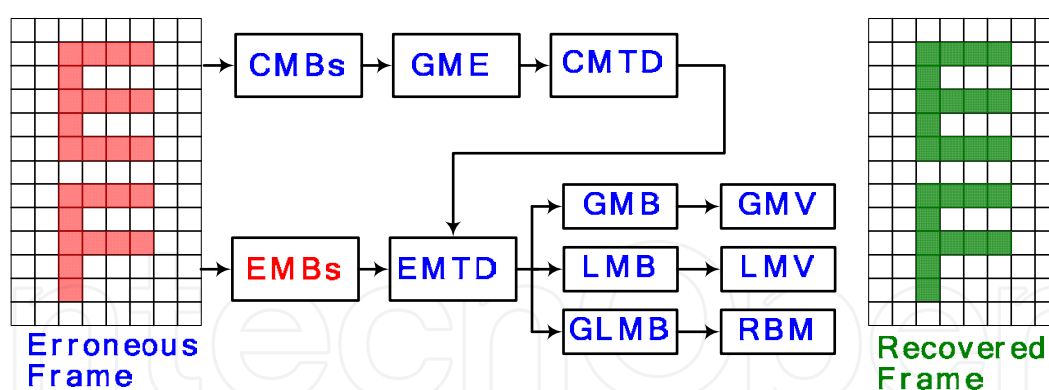

Fig. 3. Diagram of GM/LM based error concealment.

In GM/LM based EC approach, each EMB can be classified into one of the three types: GMB, LMB and GLMB according to the CMBs (including already recovered EMBs) type information in its 8-neighbors as follows:

1. If the CMBs in the neighbors of an EMB are all with the type GMB, then we classify the EMB be a GMB. The corrupted pixels in the EMB are replaced by the warped pixels in their reference frame by utilizing the GMV information.
2. If the CMBs in the neighbors of an EMB are all with the type LMB, then we classify the EMB be a LMB. The corrupted pixels in the EMB can be replaced by the MB in their

reference frame using the average MV of the non-corrupted or recovered MBs in its 8-neighbors. The GMV and LMV based replacement for the EMB are based on the facts that both global motion and local motion have certain consistence.

3. Otherwise the EMB is a GLMB. The EMB may contain both global and local motion regions. Boundaries between background and objects usually exist in the EMB. To determine the accurate boundaries, complicate boundary matching algorithms such as RBM, and AECOD [21] can be adopted. We use RBM method to search the optimal MV to recover the EMB.

5.3.3 Error concealment performance

Objective error concealment performances of the TR, TA, GM, RBM and GM/LM are given. Fig. 4 (a) and (b) show the objective averaged PSNR (peak signal to noise ratio) values of the EC methods applied to each of the P-frame of the testing sequences *flower* and *mobile* under the PER (packet error rates) 15%. From Fig.4, we find that our GM/LM based EC method gives comparatively better recovery results.

(a) flower   (b) mobile
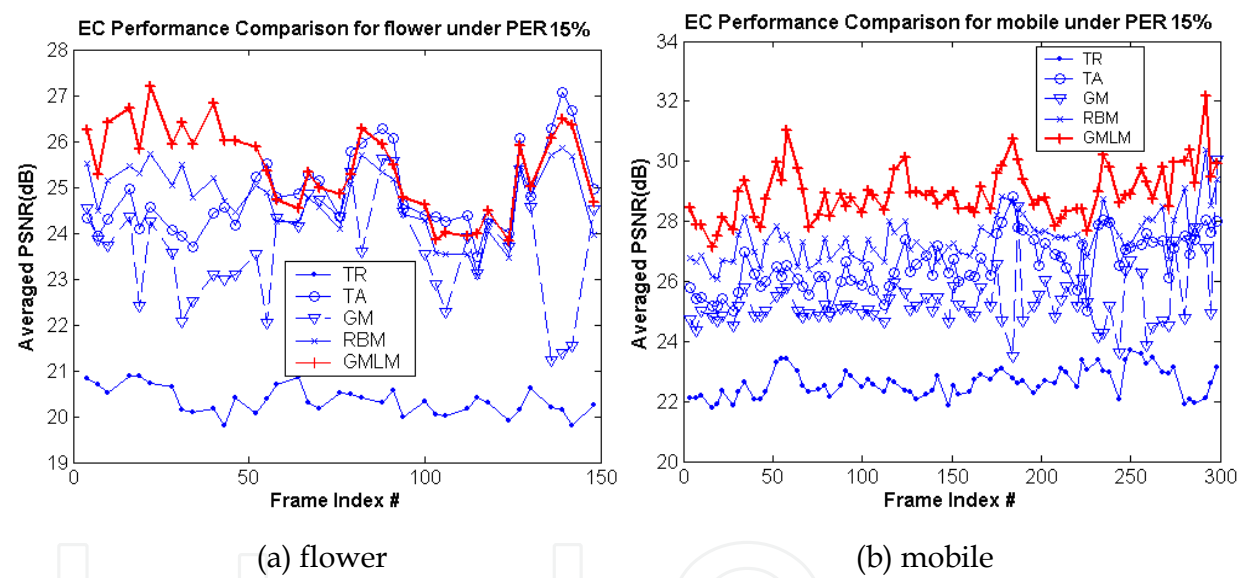
Fig. 4. EC Performances Comparison of the corresponding TR, TA, GM, RBM and GM/LM for the video sequences under PER 15%.

To show the subjective recovery results of the TR, TA, GM, RBM and GM/LM based error concealment approaches, two frames are extracted from the test video sequences with several erroneous slices, as shown in Fig. 5. We find that the recovery results of TR are not so effective. TA is not effective to get accurate motion information for the MBs in heavy motion regions. RBM performs well for the area where non-periodical texture appears. However, it is not so effective in the circumstance that the reference blocks are in smooth and texture similar regions as shown in Fig. 5(b). GM provides better recovery results for the background regions. However, large distortions are produced for recovering the EMBs in local motion regions. Comparatively, better performances are achieved by the proposed GM/LM based EC method.

a)

b)

c)

d)

e)

Fig. 5. Subjective error concealment results for #31 of flower and #28 of foreman. (a) non-error frames; (b) corrupted frames; and the corresponding recovery results of (c) TR; (d)TA;(e) GM; (f)RBM and (g) GM/LM respectively.

5.4 GM/LM based text occluded region recovery

The corresponding block diagram of the proposed GM/LM based text occluded region recovery (TORR) approach is shown in Fig.6. It consists of the following steps.

Fig. 6. Block diagram of the GM/LM based text occluded region recovery (TORR). The input video is with text occluded regions and the output video is with text occluded region recovery.
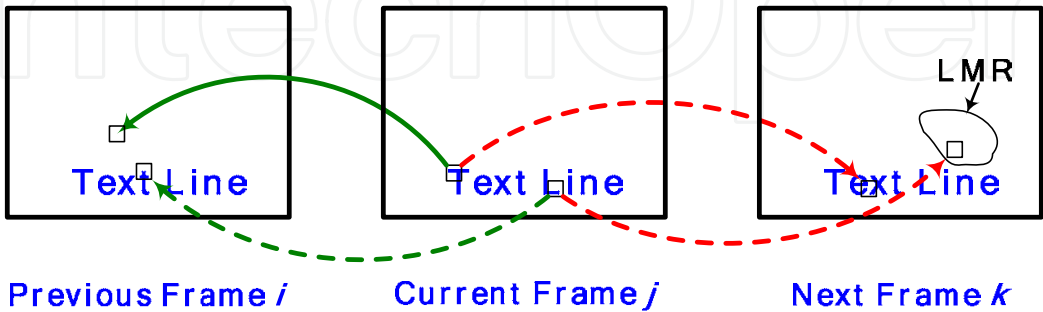
Fig. 7. Diagram of recovering a pixel in text occluded region of current frame *j* from its previous frame *i* and next frame *k*. The dash lines means the pixel cannot be recovered from its reference frames. The solid lines means the pixels can be recovered from its reference frames.

(a) video frames with detected text lines

(b) video frames after carrying out TORR

Fig. 8. GM/LM based text occluded region recovery for the three frames for the test video sequence *News1*.

(a) video frames with detected text lines

(b) video frames after carrying out TORR

Fig. 9. GM/LM based text occluded region recovery for the three frames for the test video sequence *Foxes of the Kalahari*.

1. Text detection and determining the starting/ending frames of each text line for the output video sequence. The starting and ending frames of a text line is determined by text tracking and matching. This method is with low computational costs and robust to the missing and false detections [23].
2. Video text segmentation. In the video text segmentation, text regions can be extracted by the foreground and background integrated method [23].
3. Global motion estimation. Texts are added during editing, the text regions are not undergoing global motions. So the text recovered regions can be viewed as outliers in global motion estimation.

Text occluded region recovery using GM/LM information. From the estimated global motion parameters **m** and the text occluded regions $TR = \bigcup_{i=1}^{N}(x_i, y_i)$, where N is the total pixels in the text occluded region of current frame. The corresponding diagram of a pixel in text occluded region is shown in Fig.7. TORR is carried out bi-directional and iteratively. The bi-directional approach means that a pixel in text occluded region of current frame j can be recovered by forward previous frame i and backward replacement from its next frame k (with $i<j<k$). From Fig.7 we find that the first pixel can be recovered (denoted by the solid lines) from its previous frame i and cannot recovered (denoted by the dash lines) from its next frame k. However, for the second pixel, its replacement in frame i is also in text occluded region. Moreover, its replacement in frame k is in local motion region (LMR). So the above two directional replacement are both invalid. Thus iteratively carrying out TORR is needed for the video frame. The iteration stops when all pixels in TORR are recovered. Alternatively, the replacement can be carried out by using more than one frame. It is likely that the second pixel in frame j can find correct replacement in its previous frames $i-n$ or $k+n$ (with $n>0$).

Fig.8 and Fig.9 show the subjective text occluded region recovery results. The text occluded frames in Fig.8(a) and Fig.9 (a) are from MPEG-7 test video sequences *News*1 and a documentary film of National Geography *Foxes of the Kalahari*. Fig.8 (a) and Fig.9 (a) are the video frames with detected text lines. Fig.8 (b) and Fig.9 (b) show video frames after carrying out TORR using the GM/LM based method. From the recovery results we find that the detail information of the anchorperson is kept well. This further shows the effectiveness of our GM/LM based text occluded recovery method.

6. Conclusion

In this chapter, a systematic review of the pixel domain based global motion estimation approaches is presented. With respect to its shortcomings in noise filtering and computational cost, the improvement approaches including hierarchical global motion estimation, partial pixel set based global motion estimation and compressed domain based global motion estimation are provided. Four global motion based applications including GMC/LMC in MPEG-4 video coding standard, global motion based sport video shot classification, GM/LM based error concealment and text occluded region recovery are described. The applications show the effectiveness of global motion based approaches.

7. Acknowledgement

8. Reference

[1] F. Dufaux and J. Konrad, "Efficient, robust, and fast global motion estimation for video coding", IEEE Trans. Image Processing, Vol. 9, No.3, Mar. 2000, pp.497-501.

[2] "MPEG-4 Video Verification Model version 18.0", ISO/IEC JTC1/SC29/WG11, 2001.

[3] Y. P. Su, M. T. Sun, and V. Hsu, "Global Motion Estimation From Coarsely Sampled Motion Vector Field and the Applications", IEEE Trans. Circuits Syst. Video Technol., Vol. 15, No. 2, Feb. 2005, pp. 232-242.

[4] Y. P. Su, M. T. Sun, "A Non-iterative motion vector based global motion estimation algorithm", Proc. Int. Conf. Multimedia and Expo, Taipei, Taiwan, June 27-30, 2004, pp.703-706.

[5] X. Qian, G. Liu, "Global motion estimation from randomly selected motion vector groups and GM/LM based applications", signal image and video processing, 2007, pp.179-189.

[6] Y. Keller, A. Averbuch, "Fast Gradient Methods Based on Global Motion Estimation for Video Compression", IEEE Trans. Circuits Syst. Video Technol., Vol. 13, No. 4, Apr. 2003, pp. 300-309.

[7] H. Wang, J. Wang, Q. Liu, H. Lu, Fast progressive model refinement global motion estimation algorithm with prediction, in proc. ICME 2006.

[8] H. Alzoubi, W. Pan, "very fast global motion estimation using partial data", in Proc. ICASSP 2006.

[9] H. Alzoubi, W. Pan, "efficient global motion estimation using fixed and random subsampling patterns"

[10] Y. Chen, and I. Bajic´, "Motion vector outlier rejection cascade for global motion estimation,"IEEE Signal Processing Letters, vol. 17, no. 2, 2010, pp. 197-200.

[11] M. Tok, A. Glantz, M. Arvanitidou, A. Krutz, and T. Sikora, "Compressed domain global motion estimation using the Helmholtz tradeoff estimator," in Proc. ICIP 2010,pp.777-780.

[12] Y. Chen, I. Bajic´, and P. Saeedi, "Motion region segmentation from compressed video using global motion estimation and Markov random fields," IEEE Transactions on Multimedia, vol. 13, no. 3, 2011, pp. 421-431.

[13] P. Xu, L. Xie, and S. Chang, "Algorithms and systems for segmentation and structure analysis in soccer video," in Proc. Int. Conf. Multimedia & Expo, 2001, pp. 184-187.

[14] L. Duan, M. Xu, Q. Tian, C. Xu, and J. S. Jin, "A unified framework for semantic shot classification in sports video," IEEE Trans. Multimedia, vol.7, No.6, 2005, pp.1066-1083.

[15] B. Li, J. Errico, H. Pan, and M. Sezan, "Bridging the semantic gap in sports video retrieval and summarization," J. Vis. Commun. Image R. vol.17, pp.393-424, 2004.

[16] A. Ekin, and A. Tekalp, "Generic play-break event detection for summarization and hierarchical sports video analysis," in Proc. Int. Conf. Mulmedia and Expo, vol.1, 2003, pp. 169-172.

[17] Y. Tan, D. Saur, S. Kulkarni, P. Ramadge, "Rapid estimation of camera motion from compressed video with application to video annotation," IEEE Trans. Circuits Syst. Video Techn., no.1,vol. 10, 2000,pp. 133-146.

[18] X. Qian, H. Wang, G. Liu, and X. Hou, "HMM Based Soccer Video Event Detection Using Enhanced Mid-Level Semantic", Multimedia Tools and Applications, 2011.

[19] J. Sub and Y. Ho, "Error concealment based on directional interpolation," IEEE Transactions on Consumer Electronics, vol. 43, pp. 295-302, August 1997.

[20] M. Chen, C. Chen and M. Chi, "Temporal error concealment algorithm by recursive block-matching principle", IEEE Trans. Circuits Syst. Video Technol., Vol. 15, No. 11, Nov. 2005, pp. 1385-1393.

[21] X. Qian, G. Liu, and H. Wang, "Recovering Connected Error Region Based on Adaptive Error Concealment Order Determination," IEEE Trans. Multimedia, vol.11, no.4, pp.683-695, 2009.

[22] X. Qian, and G.Liu, "An Effective GM/LM Based Video Error Concealment", Signal Image and Video Processing, 2012, vol.6,no.1, pp.9-17.

[23] X. Qian, G. Liu, H. Wang, and R. Su, "Text detection, localization and tracking in compressed videos," Signal Processing: Image Communication, vol.22 , 2007, pp.752-768.
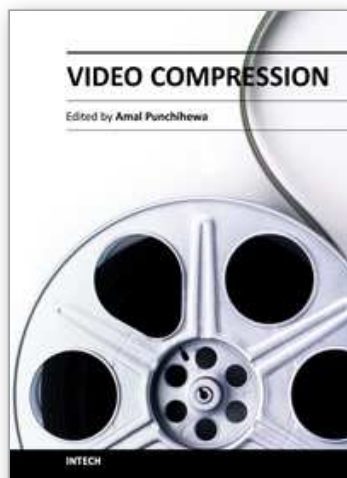
Video Compression

Edited by Dr. Amal Punchihewa

Even though video compression has become a mature field, a lot of research is still ongoing. Indeed, as the quality of the compressed video for a given size or bit rate increases, so does users' level of expectations and their intolerance to artefacts. The development of compression technology has enabled number of applications; key applications in television broadcast field. Compression technology is the basis for digital television. The "Video Compression" book was written for scientists and development engineers. The aim of the book is to showcase the state of the art in the wider field of compression beyond encoder centric approach and to appreciate the need for video quality assurance. It covers compressive video coding, distributed video coding, motion estimation and video quality.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

INTECH

open science | open minds