Australian Journal of Basic and Applied Sciences, 10(18) December 2016, Pages: 261-266

AUSTRALIAN JOURNAL OF BASIC AND APPLIED SCIENCES

ISSN:1991-8178 EISSN: 2309-8414
Journal home page: www.ajbasweb.com

Comparative Study between Motion Estimation Techniques For Video Coding

[1]Mustafa Hassan, [2]Dr. Amin Babikir, [3]Dr. Magdi Baker

[1]*Future University, Faculty of Engineering, Computer Engineering Department, Box.1055. Khartoum. Sudan.*
[2]*Anilin University, Faculty of Engineering, Telecommunication Department, Box. 12707. Khartoum. Sudan*
[3]*Gezira University, Faculty of Engineering and Technology, Electronics Engineering Department, Box.34801. Wad Madani. Sudan*

Address For Correspondence:
Mustafa Hassan, Future University, College of Engineering, Department of Computer Engineering, Lecturer. Box.1055. Khartoum. Sudan.

ARTICLE INFO

Article history:
Received 19 September 2016
Accepted 10 December 2016
Published 31 December 2016

Keywords:
Motion Estimation Techniques, Video standards, Block Matching Technique, Gradient technique, Pel-recursive Technique.

ABSTRACT

Motion estimation techniques form the core of video compression and video processing applications. As it is known video compression and processing have played important role in the world of telecommunication and multimedia system. The technique of motion estimation is extracts movement information from the video sequence where the motion is typically represented using a motion vector. The motion vector indicates the displacement of a pixel or a pixel block from the current location due to motion. This information is used in video compression to find best matching block in reference frame to calculate low energy residue to generate temporally interpolated frames. It is also used in applications such motion compensated de-interlacing, video stabilization, motion tracking etc. This paper studies and compares varieties of motion estimation techniques such as gradient , pel-recursive and block matching techniques in terms of their performance in sequence coding. The study concluded with that the block matching is the most appropriate technique to use.

INTRODUCTION

Digital video technology has become famous and known around the world over the past decades, with applications on a large scale in the field of information and communications, consumer electronics and entertainment technology. In the video sequence, the movement is the main source of information. Movement arise due to moving objects in the 3D scene, as well as the movement of the camera movement. Apparent movement, also known as optical flow, captures the resulting spatial and temporal variations of the intensities of pixels in the images in a row for the sequence. The purpose of the motion estimation techniques is to restore this information through image content analysis. Efficient and accurate motion estimation is a key element in the areas of image sequence analysis, computer vision and video communication.

In the context of image sequence analysis, computer vision and the goal of motion estimation algorithms is to precisely and faithfully model the motion in the scene. This information is fundamental for video understanding and object tracking. Relevant applications include video surveillance, robotics, autonomous vehicles navigation, human motion analysis, quality control in manufacturing, video search and retrieval, and video restoration. Accurate motion is also important in some video processing tasks such as frame rate conversion or de-interlacing.

To Cite This Article: Mustafa Hassan, Dr. Amin Babikir, Dr. Magdi Baker., Comparative Study between Motion Estimation Techniques For Video Coding. ***Aust. J. Basic & Appl. Sci.,*** *10(18): 261-266, 2016*

Australian Journal of Basic and Applied Sciences, 10(18) December 2016, Pages: 261-266

Video standards:

Basically there are two standard series in common use, both having several versions. International Telecommunications Union (ITU) started developing Recommendation H.261 in 1984, and the effort was finished in 1990 when it was approved. The standard is aimed for video conferencing and video phone services over the integrated service digital network (ISDN) with bit rate a multiple of 64 kilobits per second.

MPEG-1 is a video compression standard developed in joint operation by International Standards Organization (ISO) and International Electro-Technical Commission (IEC). The system development was started in 1988 and finished in 1990, and it was accepted as standard in 1992. MPEG-1 can be used at higher bit rates than H.261, at about 1.5 megabits per second, which is suitable for storing the compressed video stream on compact disks or for using with interactive multimedia systems (Richardson, 2003).

In 1996 a revised version of the standard, Recommendation H.263, was finalized which adopts some new techniques for compression, such as half pixel and optionally smaller block size for motion compensation. As a result it has better video quality than H.261. Recommendation H.261 divides each frame into 16×16 picture element (pixel) blocks for backward motion compensation, and H.263 can also take advantage of 8×8 pixel blocks. A new ITU standard in development is called H.26L, and it allows motion compensation.

MPEG-4: MPEG-4 became an international standard in 1998. MPEG-4 is designed to address the requirement of the interactive multimedia applications, while simultaneously supporting traditional applications. Bit rates targeted for MPEG-4 video standard range between 5-64 Kbits/s for mobile or PSTN (Public Switched Telephone Network) video applications and up to 2 Mbit/s for TV/Film applications so that this standard supersedes MPEG-1 and MPEG-2 for most applications. Video object coding is one of the most important features introduced by MPEG-4. MPEG-4 enables the possibility to manipulate and interact with the objects after they are created and compressed (Pavankumar, 2009).

Motion estimation representation:

The concept of movement should be clarified in the framework of image sequence processing. By formations in terms of velocity displacement. Due to the discrete nature of image sequences over time, instantaneous velocity and displacement of a pixel is related by a constant factor which corresponds to the temporal sampling intervals. Consequently, in this case these two quantities are interchangeable and both formulation are equivalent. However, the last statement is no longer valid when the motion of a group of pixels is modeled by a set of motion parameters. In this case, instantaneous velocity and displacement formulations may lead to distinct models. Hereafter the formulation in terms of displacement is adopted, and the term motion vector should be understood accordingly. In a digital image sequence, the 4D spatiotemporal continuum is projected onto a 3D discrete sample grid. A distinction is made between two entities: the 2D motion field and the optical flow. The 2D motion field is defined by the projection of the 3D motion on the 20 image plane. The optical flow corresponds to the spatiotemporal variation of intensity. In the ideal case. the optical flow corresponds to the 2D motion field. In video coding, motion estimation techniques estimate the trajectory of pixels over successive images in order to express the current image intensity from previous information. Therefore they estimate the optical flow. In what follows, no distinction will made between these two notions, and the term motion should be understood as optical flow(Dufaux, 2014).

In order to define a motion representation, two issues have to be considered. Firstly, a model underlying the motion representation needs to be specified. Secondly, the region of support to which the model applies has to be identified. We discuss these two subjects in more details as follow:

1- Brightness constancy motion model and the optical flow equation:

An image sequence can be considered as a three-dimensional continuous spatial temporal field: f(x,y,t), where (x,y) are the spatial coordinates and t is the time index. However, in practice we only dispose of a discrete version of this function:

$$f_{n,m,k}\ f(nl_1, ml_2, kT) \qquad (1)$$

where l_1 and l_2 are the sampling steps in the spatial domain, T is the temporal sampling step, and n, m and k are integers. In what follows, we shall make the following assumptions:

- A pixel intensity remains unchanged along a motion trajectory. This assumption is known as the brightness constancy constraint. In other words, the variations in time of the pixel intensity are due to the displacements of different objects present in the scene. The brightness constancy constraint implies that the illumination is uniform and the scene is Lumberton.
- The motion appears locally as a translation (which is the simplest, though effective, motion model).

The following is brightness constancy motion model:

$$f(x, y, t + T) = f(x - \Delta x, y - \Delta y, t) \qquad (2)$$

where Δx and Δy are, respectively, the horizontal and vertical components of the displacement. Actually, these components depend on the spatial position, the temporal index t and the temporal sampling step T.

2- Parametric motion models:

Another approach is to model the motion of a set of variables. Such a model is efficient to represent the movement of region whose pixels have cohesive motion. In addition, the parametric model result in a very compact descriptors as a only a small set of parameters to be transmitted. A simple model consists in considering in the image plane a polynomial approximation of the displacement (Tziritas and Labit, 1994). More formally, this polynomial model can be expressed as:

$$\begin{pmatrix} d_x \\ d_y \end{pmatrix} = \sum_{i,j} \begin{pmatrix} a_{i,j} \\ b_{i,j} \end{pmatrix} x^i y^j \qquad (3)$$

where dx and dy are the two components of the motion vector, ai,j and bi,j represent the parameters of the model, and (x,y) are the pixel coordinates. For example, a first-order approximation leads to a description of the form

$$\begin{aligned} d_x &= t_x + (k + h_1)x + (h_2 - a)y \\ d_y &= t_y + (h_2 + a)x + (k - h_1)y \end{aligned} \qquad (4)$$

where tx, ty, a, k, h1 and h2 are the motion model parameters. In this example, tx and ty represent the translation parameters. The parameters a, x0, y0, k, h1 and h2 are involved in the description of more complex displacements.

Region of support:

The region of support is the set of image pixels to which a motion model applies. We can distinguish the following four cases:

1- Pixel-based: A motion vector is assigned to each pixel of the image, resulting in a dense motion field. It has the advantage to provide a precise description of the motion. However, from a video coding viewpoint, it entails a costly representation resulting in a large overhead for motion information.

2- Region-based: The motion model is applied to a region of the image which is characterized by a coherent motion. In this case, moving objects in the scene have to be identified. In (Diehl., 1991), a method is presented for segmenting video scenes hierarchically into differently moving objects. A system for representing moving images with sets of overlapping layers using motion analysis in proposed in (Wang and Adelson, 1994). Segmentation-based motion estimation and spatial-temporal segmentation are addressed in (Dufaux and Moscheni., 1996). In (Chang and Sezan, 1997) a Bayesian framework is presented that combines motion estimation and segmentation. In the context of video coding, such a representation requires to transmit the shape of the region, which entails a bit rate overhead.

3- Block-based: As a special case of the region-based support, a very frequent choice is to simply partition the image into blocks. If the block size is sufficiently small, then the assumption that the block is moving in a coherent way is likely to be valid. Another advantage of a block partitioning is that it does not require additional information to represent the shape of the region.

4- Global: The region of support simply encompasses the whole image. This case is especially suited to efficiently estimate camera motion. Indeed, camera motion such as dolly, track, boom, pan, tilt or roll, is an essential cinematic technique.

The choices of the region of support and the motion model are closely intertwined. When using a complex parametric motion model, which can handle complex motions, a larger region of support can effectively be used. Conversely, a simple model is often sufficient in conjunction with a small region of support (Dufuax and et al., 2014).

Motion estimation techniques:

Motion estimation techniques can be divided into four main groups:

1- Gradient techniques
2- Pel-recursive techniques
3- Block matching techniques
4- Frequency-domain techniques

The gradient technologies have been developed for applications of image sequence analysis. They solve visual flow and results in a dense motion field. Both techniques pel-recursive and a block matching has been developed within the framework of image sequence coding. The techniques pel- recursive can be considered as a subset of gradient techniques. However as they constitute an important contribution in the field of coding, we consider them as a separate group. Block matching techniques are based on the minimization of a disparity measure. They are the most widely used in coding applications. Finally. frequency-domain techniques are based on the relationship between transformed coefficients (e.g., Fourier or Gabor transform) of shifted images. However they lack a widespread use, especially in the field of image sequence coding.

In the following, the image intensity at pixel location $\vec{r} = (x,y)^T$ and at time t is denoted by: $I(\vec{r}, t)$ The displacement during the interval Δt. All the techniques presented here after rely on the hypothesis that a change in the image intensity $I(\vec{r}, t)$ is due only to the displacement $\vec{d}$ .It is expressed by:

$I(\vec{r}, t) = I(\vec{r} - \vec{d}, t - \Delta t)$ – (5)

And the displaced frame difference (DFD) is defined as:

$$DFD(\vec{r}, t, \vec{d}) = I(\vec{r}, t) - I(\vec{r} - \vec{d}, t - \Delta t) \quad (6)$$

1- Gradient Techniques:

Gradient techniques rely on the hypothesis that the image luminance is invariant along motion trajectories. The Taylor series expansion of the right hand side in (5) gives:

$$I(\vec{r}, t) - I(\vec{r} - \vec{d}, t - \Delta t) = I(\vec{r}, t) - \vec{d}.\vec{\nabla} I(\vec{r},t) - \Delta t \frac{\partial I(\vec{r},t)}{\partial t} + Higher\ order\ terms \quad (7)$$

Where $\vec{\nabla} = \left\{\left(\frac{\partial}{\partial x}\right), \left(\frac{\partial}{\partial y}\right)\right\}^T$ is the gradient operator. Neglecting the higher order terms (first order approximation), assuming the limit $\Delta t \rightarrow 0$ and defining motion vector $\vec{v} = (v_x, v_y)^T = \frac{\vec{d}}{\Delta t}$, we obtain:

$$\vec{v}.\vec{\nabla} I(\vec{r},t) + \frac{\partial I(\vec{r},t)}{\partial t} = 0 \quad (8)$$

The latter equation is known as the spatial-temporal constraint equation or the optical flow constraint equation. As the image intensity change at a point due to motion gives only one constraint (8), while the motion vector at the same point has two components, the motion field (actually the optical flow) cannot be computed without an additional constraint. In fact, only the projection of $\vec{v}$. on $\vec{\nabla} I$. in other words the component of $\vec{v}$. parallel to the intensity gradient, can be determined from (8). This problem is known as the aperture problem. Therefore an additional constraint must be introduced to regularize the ill-posed problem and to solve the optical flow. In Hom Schunck introduce a smoothness constraint, that is to minimize the square of the optical flow gradient magnitude:

$$\left(\frac{\partial v_x}{\partial x}\right)^2 + \left(\frac{\partial v_x}{\partial y}\right)^2 \text{ and } \left(\frac{\partial v_y}{\partial x}\right)^2 + \left(\frac{\partial v_y}{\partial y}\right)^2 \quad (9)$$

Consequently, the optical flow is obtained by minimizing the following error term

$$\left\{\iint \left\{\left(\vec{v}.\vec{\nabla} I + \frac{\partial I}{\partial t}\right)^2 \propto^2 \left[\left(\frac{\partial v_x}{\partial x}\right)^2 + \left(\frac{\partial v_x}{\partial y}\right)^2 + \left(\frac{\partial v_y}{\partial x}\right)^2 + \left(\frac{\partial v_y}{\partial y}\right)^2\right]\right\} dx\, dy \quad (10)$$

Where $\propto^2$ is a weighting factor. This minimization problem is solved by the variation calculus and an iterative Gauss-Seidel procedure. Many variations of the above algorithm have been proposed. Instead of the smoothness constraint, an assumption of local uniformity is made by Lucas. Nagel develops the Taylor series of (7) up to the second-order terms. Nagel introduces an oriented smoothness constraint, which takes occluding edges into consideration, instead of the isotropic smoothness constraint used in (10). All these techniques result in a dense motion field (Le., a motion vector per pixel). This is qualitatively interesting for motion analysis applications. However, from an image sequence coding point of view, these techniques suffer from two serious drawbacks. First the smoothness constraint leads to an increased energy of the prediction error, especially on moving objects boundaries. Second, the dense motion field requires much overhead information. This problem is tackled through the use of a motion vector field coded at different resolutions. At each level, only the motion vectors from which the motion compensation benefits most are coded (Dufaux, 1995).

2-Pel-Recursive Techniques:

Pel-recursive techniques recursively minimize the prediction error or DFD in (6). The recursion is usually carried out on a pel-by-pel basis, leading to a dense motion vector field. These methods are among the first motion estimation techniques developed for image sequence coding applications and have been designed with a constraint of a very low hardware complexity. Pel-recursive techniques can be considered as a subset of the gradient techniques in which the spatiotemporal constraint is minimized recursively. The first pel-recursive algorithm was proposed by Netravali Robbins. In this algorithm, the DFD2 is iteratively minimized by the steepest descent technique,

$$\vec{d}^{(k+1)} = \vec{d}^{(k)} - \frac{\in}{2} \vec{d} DFD^2\left(\vec{r}, t, \vec{d}^{(k)}\right) \quad (11)$$

with a constant gain $\in > 0$, and k denotes the iteration index. From the definition of the DFD, (6), we have

$$\nabla_{\vec{d}} DFD^2(\vec{r},t,\vec{d}) = 2DFD(\vec{r},t,\vec{d}).\nabla_{\vec{r}} I(\vec{r}-\vec{d},t-\Delta t) \tag{12}$$

By substituting (12) in (11) the displacement vector update becomes

$$\vec{d}^{(k+1)} = \vec{d}^{(k)} - \in DFD\left(\vec{r},t,\vec{d}^{(k)}\right).\nabla_{\vec{r}} I(\vec{r}-\vec{d}^{(k)},t-\Delta t) \tag{13}$$

The performance of the pel-recursive techniques depends strongly on the way to compute the update term in (13). In the algorithm introduced in r 61. the iteration from k to k+ 1 is carried out either on one pel location, or from one pel to its consecutive neighbor. To smooth out the effect of noise, the algorithm can be extended by calculating and averaging the update term on several pels. Improved algorithms based on the same principle have been proposed for instance by Cafforio and Rocca, and Walker and Rao (Dufaux, 1995). Compared to (6) the gain $\in$ is substituted by variables in order to achieve a better adaptation to the local image statistic and consequently a better convergence. Special care has been taken to enhance the tracking capability and the stability of the pel-recursive algorithms. A multiple frames model-based approach is presented, whereas also a multiple mask regularization technique is proposed. The algorithm introduced in (M.Orchard, 1992) shares characteristics of both pel-recursive and block matching approaches by combining motion information sent from the encoder and motion information recursively estimated at the decoder. Provided that the recursion has a sufficiently rapid convergence (i.e., it can handle motion discontinuities), such pel-recursive algorithms may overcome the problem of multiple moving objects. Furthermore, when the update of the displacement vector is based only on previously transmitted data (causality), the decoder is able to estimate the same motion field than the encoder. In this case, no overhead motion information is required, which is of course a further advantage of these methods. However, the causality constrains these algorithms and reduces their prediction capability relative to non causal methods. Furthermore, the pel-recursive motion estimation technique (with recursion on pels) is not compatible with transform coding of the DFD, as in this case the decoder is unable to reconstruct the motion vectors. Moreover, the algorithm implies an increased computational complexity at the decoder, as the latter should also estimate the motion field. Pel-recursive algorithms suffer from two further drawbacks:

First, as the error function to be minimized contains generally many local minima, the iterative procedure may converge to a local minimum rather than to the global one. In particular, these algorithms are very sensitive to noise.

Second, large displacements and discontinuities in the motion field cannot be efficiently handled.

3-Block Matching Techniques:

Block matching algorithms are based on the matching of blocks between two images, the aim being to minimize a disparity measure. Specifically developed for image sequence coding, they are widely used in this field. In block matching motion estimation, the image is partitioned into blocks and the same displacement vector is assigned to all pixels within a block. The motion model usually assumes that an image is composed of rigid objects in translational motion. Although this model is clearly restrictive, it is justified by the fact that complex motion can be decomposed as a sum of translational components. Consider the problem of predictive coding, the aim of motion estimation is to find the displacement vector $\vec{d},$ which allows predicting I($\vec{r},t$) from I($\vec{r},t$-Δt) in (5). For each block, the displacement vector is evaluated by matching the information content of a measurement window *W* with that of a corresponding measurement window within a search area S, placed in the previous frame, and by searching the spatial location minimizing the matching criterion:

$$\vec{d} = argmin_{\vec{d} \in s} \sum_{\vec{r} \in w} ||I(\vec{r},t) - I(\vec{r}-\vec{d},t-\Delta t)|| \tag{14}$$

where the most widely used distance measures are the quadratic normal ||x|| = x^2 and the absolute value ||x|| = |x|. Finding an absolute minimum for the matching criterion can only be guaranteed by performing an exhaustive search of a series of discrete candidate displacements within a maximum displacement range, this technique is called full-search block matching. Despite the heavy computations it requires, it is widely used in video coding, due to its simplicity and ease of hardware implementation. In order to decrease the computational load of the full search algorithm, fast search techniques have been proposed. These fast search techniques afford significantly reduced computation times compared to the full-search algorithm. However, convergence toward the global minimum is guaranteed only when the matching criterion is a monotonic function of $\vec{d}$. Block matching algorithms have been designed initially to estimate displacements with a precision of one pixel, however a sub-pixel accuracy can be obtained. For this purpose, the image intensity has to be interpolated at fractional pixel locations. In practice, this stage is implemented in post-processing where the one pixel accuracy

Australian Journal of Basic and Applied Sciences, 10(18) December 2016, Pages: 261-266

displacement vectors are refined to a fractional pixel precision. Notably, this post-processing significantly increases the computational complexity.

To conclude, as the block matching methods directly minimize the DFD, they are well suited for image sequence coding. Furthermore, due to the block-based nature of these techniques, they require only little overhead motion information. For these reasons, block matching motion estimation techniques are the most widely used in image sequence coding. Recent standards such as MPEG - l, MPEG-2 and H.261 are based on them, even though the algorithm to estimate the motion vectors is not specified explicitly.

Comparisons between techniques:

To evaluate the performances of the different motion estimation techniques, the following comparisons were made:

First, it was rated the subjective quality of the field estimated movement, showing the algorithm's ability to estimate the true movement in the scene. In particular, the required areas of smooth motion in the coding to prevent interruptions in artificial DFD and the reduction of public expenditures for the transfer of information movement.

Second, the energy measure DFD error, giving an idea of the quality of the prediction, which is a major feature in the coding. Who can measure each technique as follows:

In the Hom-Schunck gradient technique, obtaining a dense motion field does not lead to an improved prediction capability, whereas it does induce a high amount of overhead information. Therefore, the method is more interesting from an analysis rather than coding point of view.

1. In the Netravali-Robbins pel-recursive technique, even though it provides a dense motion field, performs poorly in estimating the motion. The savings in overhead information is not sufficient to compensate for this poor prediction.
2. In the Block matching technique in spite of its drawbacks, which directly minimizes the DFD energy and requires only one motion vector per block, appears to be the most suitable for coding purposes. It relies on a simple motion model which leads simultaneously to a precise motion estimation and a low overhead. Therefore, it achieves a good allocation of the bandwidth between DFD and motion parameters. Furthermore, in the current standards MPEG-l, MPEG-2 and H.261 which are based on a DCT transform coding, block matching motion estimation techniques are clearly the most appropriate.

Conclusion:

The paper is reviewed and explained the advantages and disadvantages of motion estimation techniques in terms of their performances in sequence coding. The comparative analysis put two parameters for evaluating the performance of the techniques. The block matching technique is appeared that it is the most one suitable for coding purpose. This review study will make a clear picture to the researchers that are working in this field, either for applying or developing techniques.

REFERENCES

Lucas, B. and T. Kanad., 1981. An iterative Image Registration Technique with An application to Stereo Vision. In the Proceedings of the 1981 Imaging Understanding Workshop, pp: 121-130.

Chang, M.M., A.M. Tekalp and M.I. Sezan, 1997. Simultaneous Motion Estimation and Segmentation. Journal of IEEE Transactions on Image Processing, 6(9): 1326-1333.

Diehl, N., 1991. Object-oriented Motion Estimation and Segmentation in Image Sequences. Journal of Signal Processing: Image Communication, 3(1): 23-56.

Dufaux, F. and F. Moscheni, 1996. Segmentation-based Motion Estimation for Second Generation Video Coding Techniques, in Second Generation Video Coding Techniques, L. Torres, Kunt Kluwer Academic Publishers, pp: 219-263.

Frédéric Dufaux, Marco Cagnazzo, B. Pesquet-Popescu., 2014. Motion Estimation Techniques. Telecom Paris Technology Press.

Nagel, H.-H., 1983. Displacement Vectors Derived from Second-order Iintensity Variations in Image Sequences. Journal of Computer Graphics and Image Process, 21(1): 85-1 17.

Richardson, I.E.G., 2003. H.264 and MPEG-4 Video Compression. Wiley Chichester Press.

Orchard., M., 1992. New Pel-recursive Motion Estimation Algorithms Based on Novel Interpolation kernels. In proceeding of 1992 Visual Communication and Image Process Conference, pp: 85-96.

Pavankumar Gorpuni., 2009. Development of Fast Motion Estimation Algorithms for Video Compression, M Tech thesis, National Institute of Technology., Rourkela, India.

Tziritas, G and C. Labit, 1994. Motion Analysis for Image Sequence Coding. Elsevier Science Press.

Wang, J.Y.A and E.H. Adelson, 1994. Representing Moving Images with layers. Journal of IEEE Transactions on Image Processing, 3(5): 625-638.