AN ADAPTIVE MOTION ESTIMATION SCHEME USING MAXIMUM MUTUAL INFORMATION CRITERIA

Jing Zhao†, Deniz Erdogmus§, Cancan Huang†, Dapeng Wu†, Yuguang Fang†

†Department of Electrical & Computer Engineering, University of Florida
Gainesville, FL 32611-6130,USA. web: www.wu.ece.ufl.edu
§Department of CSEE, Oregon Health and Science University, 20000 NW Walker Road, Beaverton, OR 97006, USA
phone: + (1) 503-7482007, fax: + (1) 503-7481548, email: derdogmus@ieee.org

ABSTRACT

Motion estimation in video coding can be formulated as an optimization problem. Recently, a motion estimation scheme that uses Renyi's error entropy as the optimization criterion, was proposed [1]. Motivated by [1], in this paper, we propose a different criterion in motion estimation, i.e., the criterion of maximum mutual information. Based on this new criterion, we design a motion estimation algorithm. Our results show that our algorithm achieves significantly lower computational complexity compared to existing fast-search methods for motion estimation. A salient feature of our algorithm is that it is ideally suited for wireless video sensor networks where limited bandwidth, restricted computational capability, and limited battery power supply pose stringent constraints on the system.

1. INTRODUCTION

Recently there has been a considerable increase in manufacturing and use of mobile communication devices equipped with video cameras. The last several years see a surging interest in transmission of video over wireless network. Many of the mobile communication devices are small and battery operated. Therefore they have only a limited amount of power and low computation capability. This pushes the needs for more efficient video compression algorithms.

To achieve coding efficiency, intra-frame coding and inter-frame coding are utilized to reduce spatial redundancy within a single frame and temporal redundancy between adjacent frames. As a key component of most video compression systems, motion estimation exploits the temporal redundancy by predicting the subsequent frames from reference frames. Motion estimation constitutes 70% of the computation load in encoder [3]. Thus for resource-constrained wireless video applications, there is an urgent need for motion estimation scheme with low computation complexity.

Block-based motion estimation schemes are the most widely used technique. In schemes of this category, each video frame is divided into blocks. All the pixels in the block are assumed to undergo the same translational motion specified by the motion vector of this block. The motion vector is estimated by searching for the best matching block within a search window centered on the corresponding block in the reference frame. Thus each block can be predicted from the previously coded reference frame based on the motion vector of the block. And the prediction error is coded with intra-frame coding techniques.

In wireless video applications, an exhaustive search may be neither realistic due to its formidable computation complexity, nor cost-effective as the motion is not completely random. Many algorithms were developed to perform motion estimation with reduced computational complexity, among which are two-dimensional logarithmic (TDL) search [5], block-based gradient descent search [6], three-step search [7], a new three-step search (TSS) [8], the four-step (4SS) search [10], to name a few.

But all the block-based motion estimation algorithms mentioned above do not effectively utilize the knowledge gained in calculating the motion vectors from one frame to the next. For each search, they usually reset their memory and start from the same initial conditions.

One way to fully utilize the information gained from the past frames for the estimation of future frames is to formulate the motion estimation problem as an adaptive filtering problem. In such a video compression system, motion vectors are modeled by an adaptive system, in contrast to no modeling of motion vectors in traditional approaches. With this formulation, we present in this paper a new approach to determine the motion vector in information-theoretic frameworks. The advantage of our scheme lies mainly in the extremely low computation complexity it achieves. Furthermore, since the motion vector model is to be replicated at the decoder given knowledge about the initial conditions, there is no need to transmit motion vectors! This provides savings in bandwidth on top of the saving in computation.

The remainder of the paper is organized as follows. In Section 2, we introduce adaptive systems used for motion estimation. In Section 3, we derive our adaptive motion estimation algorithm using a maximum mutual information criterion and analyze the computational complexity of the scheme. In Section 4, we present the simulation results of our algorithm and compare it with the existing schemes in terms of root mean squared error (RMSE). Section 5 concludes the paper.

2. MOTION ESTIMATION PROBLEM IN AN ADAPTIVE SYSTEM FRAMEWORK

Adaptive filters have been successfully used in various research areas including signal processing, telecommunication, system identification, and automated control. In this section, we formulate the motion estimation problem in an adaptive system framework.

Let $f(p,n)$ denote the image intensity at spatio-temporal position (p,n), where $p=[x,y]$ is the pixel location in 2-dimensional space, n is the time index. Given two successive frames $f(n)$ and $f(n+1)$, a motion vector (MV) $d(n)=[d_x(n),d_y(n)]$ is defined for each pixel as the 2-D vector field that maps the point in $f(n)$ onto their corresponding location in $f(n+1)$. Our goal is to find an estimate of d based

on values of $f(n)$ and $f(n+1)$, so that some pre-defined cost function J is optimized.

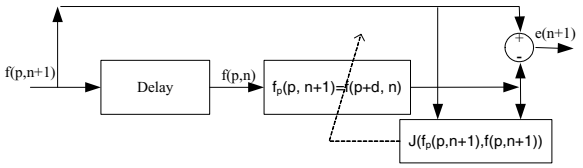

Figure 1: Adaptive Motion Estimation System Diagram

The block diagram of a general adaptive prediction system is shown in Fig. 1. The filter input is the intensity function of the most recent frame, $f(p,n)$. For the adaptive filter, the desired output is the intensity function of the present frame $f(p,n+1)$. At some discrete time, n+1, the output of the filter, $f_p(p,n+1)$, is the estimate of the $f(p,n+1)$ given its most recent values $f(p,n)$. The estimation error, $e(n+1)$ is defined as the difference between the filter output $f_p(p,n+1)$, and the desired output $f(p,n+1)$. $J(f_p(p,n+1),f(p,n+1))$ is the optimization criterion, which is a function of $f_p(p,n+1)$ and $f(p,n+1)$.

The equation for the motion estimation system described above can be expressed as follows

$$f_p(p,n+1) = f(p+d(n),n) \tag{1}$$

$$e(p,n+1) = f(p,n+1) - f_p(p,n+1) \tag{2}$$

$$f_r(p,n+1) = f_p(p,n+1) + e(p,n+1) \tag{3}$$

$$d(n+1) = d(n) + \frac{J}{d(n)} \tag{4}$$

where $f_r(p,n+1)$ is the reconstructed value for the pixel at location p in frame $n+1$, and is the updating step size. For the same initial conditions and error-free transmission over the channel, both encoder and decoder can obtain the same motion vectors in real-time, thereby avoiding the need for transmitting them to the decoder. So, only error signal e will be transmitted. In reality, due to bandwidth restriction, the error signal is quantized and it leads to error accumulation in image reconstruction. This problem can be solved by introducing intra frames periodically. Note that in this paper, our discussion is focused on motion estimation under the assumption that lossless transmission is used for error signal.

Given the system equations, if the cost function takes the form of mean square error (MSE) and is represented as $J(e)$, we get the popular LMS algorithm. Recently, Renyi's error entropy has been proposed as an alternative optimization criterion for system modeling [2]. Based on this criterion, the minimum error entropy motion estimation scheme was developed in [1] to achieve low computation complexity.

Motivated by all these works, we use another information theoretic criterion, i.e., mutual information, to develop our scheme. The intuition behind this is when the mutual information between the predicted signal $f_p(p,n+1)$ and real signal $f(p,n+1)$ is maximized, the predicted frame will preserve the most information about the real frame, thus obtain the optimal prediction. However, there are no analytical methods to calculate mutual information without presuming knowledge of prior probability density function (pdf). To avoid this, we use a non-parametric mutual information estimator with Parzen Windowing, which can be applied directly to data samples without imposing any assumptions about the pdf of the data. Thus the method can manipulate information as straightforwardly as the mean square error (MSE) criterion. Straightforward methods that use the Quadratic Renyi's entropy in a way similar to LMS method have been developed in [2]. We follow this approach to develop our motion estimation scheme in the next section.

3. MOTION ESTIMATION SCHEME BY MAXIMIZING MUTUAL INFORMATION

This section is organized as follows. First, a brief review of non-parametric pdf estimator with Parzen windowing is given. It is followed by the proposal of mutual information as the cost function for the motion estimation system. Following this is the derivation of the stochastic gradient estimator of the cost function with respective to the motion vector. Finally, we give the complexity analysis.

3.1 Non-parametric pdf estimator

Given samples of random variables x and y, one way to estimate the data pdf lies in the use of Parzen Window method. For 2-D random vector $z = (x,y)^T$, given N pairs of samples, the pdf can be approximated by Parzen windowing estimator with Gaussian kernel with zero mean and covariance matrix ,

$$p_{X,Y}(x,y) = \frac{1}{N}\sum_{i=1}^{N} G\ (x-x_i, y-y_i) \tag{5}$$

where G is the Gaussian kernel with zero mean and covariance matrix in Parzen windowing.

Similarly, the marginal pdf of random variable x, y can be approximated as

$$p_X(x) = \frac{1}{N}\sum_{i=1}^{N} G\ _{\frac{2}{X}}(x-x_i) \tag{6}$$

$$p_Y(y) = \frac{1}{N}\sum_{i=1}^{N} G\ _{\frac{2}{Y}}(y-y_i) \tag{7}$$

where $^2_X = \ _{11}$ and $^2_Y = \ _{22}$ are the kernel variance of x and y respectively.

Thus the pdf estimators solely based on the data without assuming any *a prior* knowledge of the distribution of the data are obtained. This estimator is used in the development of stochastic gradient estimator.

3.2 Mutual information as a cost function

Mutual information is widely used as a measurement of the similarities or discrepancies. According to Shannon's definition of mutual information, for two random variable x and y,

$$\mathrm{I}_s(x;y) = \int\int p_{X,Y}(x,y)\log\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}dxdy \tag{8}$$

where $p_{X,Y}(x,y)$ is the joint pdf of x,y, and $p_X(x)$ and $p_Y(y)$ are the marginal pdf's of x,y respectively. Equation (8) can be simplified as

$$\mathrm{I}_s(x;y) = \mathscr{E}(\log\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}) \tag{9}$$

In image processing, the maximum mutual information criterion has been used successfully to solve the problem of image registration. In the motion estimation problem, it seems to us that the mutual information between the system output $f_p(p,n+1)$ and the desired output $f(p,n+1)$ is a natural criterion, thus by maximizing the mutual information $I_s(f_p(p,n+1), f(p,n+1))$, the optimal motion vector d^* could be obtained.

3.3 Stochastic gradient estimator of mutual information

In [2], a stochastic gradient estimator was developed by applying the complexity reduction techniques to Renyi's entropy of error signal. For random variable x and y, when only two pairs of samples of $(x,y)^T$ are available, the non-parametric pdf estimator could be obtained as in Section 3.1. And by applying the same technique as in [2], a non-parametric stochastic estimator for mutual information is obtained:

$$I_s(x;y) \approx \log \frac{G\ (x_j - x_i, y_j - y_i)}{G_{\ 11}(x_j - x_i) G_{\ 22}(y_j - y_i)} \tag{10}$$

By far we have obtained a non-parametric estimator of $I_s(x;y)$, the mutual information between x and y. As described in Section 3.2, to develop an online adaptation algorithm for the adaptive system shown in Fig.1, we take

$$J(f_p(p,n+1), f(p,n+1)) = I_s(f_p(p,n+1); f(p,n+1)) \tag{11}$$

Substituting $f_p(p,n+1)$ and $f(p,n+1)$ for x and y in Equation (10), we obtain the cost function based on the most recent frames only:

$$\begin{aligned} &J(f_p(p,n+1), f(p,n+1)) \\ &= \log \frac{G\ (f_p(p,n+1) - f_p(p,n), f(p,n+1) - f(p,n))}{G_{\ 11}(f_p(p,n+1) - f_p(p,n)) G_{\ 22}(f(p,n+1) - f(p,n))} \end{aligned} \tag{12}$$

where $f_p(p,n+1)$ and $f_p(p,n)$ are obtained from Equation (1) by mapping the pixel on the frame $f(p,n)$ and $f(p,n-1)$ to $f_p(n+1)$ and $f_p(n)$ knowing motion vector $d(p,n)$ and $d(p,n-1)$, respectively.

The parameter to be determined is $d(p,n)$. Note that here $d(p,n-1)$ is a known constant at this time, since it is the displacement from the previous frame.

Let
$K_1 = G\ (f_p(p,n+1) - f_p(p,n), f(p,n+1) - f(p,n))$,
$K_2 = G_{\ 11}(f_p(p,n+1) - f_p(p,n))$,
$K_3 = G_{\ 22}(f(p,n+1) - f(p,n))$.
By observing Equation (12), we find that among these three terms, only K_1 and K_2 depend on $d(p,n)$. Thus we derive the partial derivatives with respect to $d(p,n)$ as

$$\frac{J}{d(p,n)} = \frac{\frac{K_1}{d(p,n)}}{K_1} - \frac{\frac{K_2}{d(p,n)}}{K_2} \tag{13}$$

where

$$\begin{aligned} &\frac{K_1}{d(p,n)} = \\ &-K_1 \cdot ((\ _{12}' + \ _{21}')(f(p,n+1) - f(p,n)) \\ &+2\ _{22}'(f(p + d(p,n), n) - f(p + d(p,n-1), n-1))) \\ &\begin{bmatrix} (f(p+e_1,n) - f(p-e_1,n))/2 \\ (f(p+e_2,n) - f(p-e_2,n))/2 \end{bmatrix} \end{aligned} \tag{14}$$

and

$$\begin{aligned} &\frac{K_2}{d(p,n)} = \\ &-\frac{1}{\ _{11}} K_2 \cdot (f(p + d(p,n), n) - f(p + d(p,n-1), n-1)) \\ &\begin{bmatrix} (f(p+e_1,n) - f(p-e_1,n))/2 \\ (f(p+e_2,n) - f(p-e_2,n))/2 \end{bmatrix} \end{aligned} \tag{15}$$

where $\ ' = \ ^{-1}$, the inverse matrix of $\ $, $e_1 = [1,0]^T$, $e_2 = [0,1]^T$. Thus we obtain a simple expression of the stochastic gradient:

$$\begin{aligned} &\frac{J}{d(p,n)} = \\ &(-\frac{1}{\ _{11}'}(f(p + d(p,n), n) - f(p + d(p,n-1), n-1)) \\ &-((\ _{12}' + \ _{21}')(f(p,n+1) - f(p,n)) \\ &+2\ _{22}'(f(p + d(p,n), n) - f(p + d(p,n-1), n-1)))) \\ &\begin{bmatrix} (f(p+e_1,n) - f(p-e_1,n))/2 \\ (f(p+e_2,n) - f(p-e_2,n))/2 \end{bmatrix} \end{aligned} \tag{16}$$

So far we have obtained an stochastic gradient estimator for the cost function J with respect to motion vector $d(p,n)$ by following a methodology similar to the minimum Renyi's error entropy algorithm described in [2].

Finally, by substituting Equation (16) in Equations (1) to (4), and imposing a smoothness constraint that the neighboring motion vectors cannot differ by more than a predetermined threshold, we obtain the Maximum Mutual Information (MaxMI) motion estimation scheme.

3.4 Computational complexity analysis

By examining Equation (4) and Equation (16), we find that on the encoder side, each pixel takes 15 operations for one frame. Table 1 shows the number of operations of the encoder and decoder of different schemes in the worst case, assuming that the search range is 16x16 and block size is 3x3. The minimum error entropy scheme(MinEE) proposed in [1] is also listed below.

Table 1: Number of instructions per pixel.

Method	Encoder	Decoder
EBMA	3267	2
Block-based Gradient Descent	252	2
Three Step Search	99	2
MinEE	15	15
MaxMI	15	15

This results shows that compared to Exhaustive Block Matching Algorithm (EBMA), Block-based Gradient Descent, and TSS, MaxMI and MinEE algorithms achieve extremely low computation complexity on the encoder, and much higher computation complexity on the decoder. This makes it ideal for the applications where the encoders are resource-constrained, e.g., wireless video sensors and mobile phones, while the decoders are more sophisticated and not much constrained by power supply, e.g., base stations.

4. SIMULATION RESULTS

In this section, we implement our adaptive motion estimation algorithm as described in Section 3. We choose the luminance component of several video sequences in QCIF format

for the encoding process. For EBMA, a block size of 8x8 is chosen with integer-pel accuracy. The search range is 16x16 pixels. The block-based gradient descent search algorithm is implemented as described in [6] with a block size of 3x3 and a search range of 16x16 pixels with integer-pel accuracy. For the three-step algorithm [7], we use a block size of 8x8 and a search range of 16x16 pixels with integer-pel accuracy. The mean absolute error (MAE) distortion function is used as the block distortion measure for the two algorithms. Since we focus on the study of motion estimation, hence DCT, quantization and entropy coding are excluded in the simulation.

In each algorithm, motion is estimated and compensated using perfectly reconstructed reference frames. The first frame is intra-coded and the rest, inter-coded. The experiment is conducted using frame rates of 10, 5 and 2, respectively. The values of root mean squared error (RMSE) for the four different QCIF sequences are shown in Tables 2, 3 and 4. The preliminary results show that the RMSE of our

Table 2: RMSE for 4 test video sequences at 10 fps

Method	Miss America	Coastguard	Suzie	Foreman
EBMA	2.88	9.29	4.92	8.16
Three-step	4.06	12.14	9.57	16.24
Gradient Descent	6.78	14.27	16.28	23.69
MinEE	6.19	20.66	12.49	20.92
MaxMI	6.07	21.03	12.08	20.73

Table 3: RMSE for 4 test video sequences at 5 fps

Method	Miss America	Coastguard	Suzie	Foreman
EBMA	3.16	11.18	6.35	11.00
Three-step	5.28	11.94	12.93	21.96
Gradient Descent	8.51	18.03	19.32	28.55
MinEE	8.85	24.95	17.68	29.14
MaxMI	8.79	23.01	20.16	29.98

Table 4: RMSE for 4 test video sequences at 2 fps

EBMA	3.63	14.29	8.69	17.91
Three-step	9.71	23.39	17.99	31.92
Gradient Descent	12.40	22.18	24.10	37.76
MinEE	13.06	29.75	24.92	39.59
MaxMI	14.42	27.75	22.02	36.28

algorithm is larger than that of the three-step search. However, our scheme does not require the transmission of motion vectors, which usually constitutes about 50% of the total bit budget for low bit-rate video applications, leading to saving of bandwidth.

5. CONCLUSIONS

Motion estimation is a critical problem in the design of a video encoder. In this paper, we proposed to use maximum mutual information as an optimization criterion to solve the motion estimation problem in the framework of adaptive system and derived a completely new scheme with low computation complexity. In this scheme, the motion vectors of the current frame are iteratively computed from the previous frame, resulting in computational savings because of the knowledge gained in the computation of the previous motion vectors. And because the motion vectors are generated automatically on the decoder side and need not be transmitted, bandwidth savings is significant. Our results showed that our scheme reduces the computational complexity on the encoder side significantly as compared to the existing fast algorithms.

The nice feature of adaptive motion estimation algorithm is its very low computational complexity, which makes it ideally suited for wireless video applications, in which computational complexity and energy consumption pose major constraints on the system. With the emergence of wireless video sensor networks, we expect that our algorithm will find widespread applications.

REFERENCES

[1] G. Ramachandran, V. Krishnan, D. Wu, Z. He, "Very low complexity, model-based motion estimation using renyi entropy for wireless video," to appear in *Journal of Visual Communication and Image Representation*.

[2] D. Erdogmus, J. Principe, "Comparison of entropy and mean square error criteria in adaptive system training using higher order," in *Proc. Intl. Workshop on Independent Component Analysis and Signal Separation 2000*, Helsinki, Finland, 2000, pp. 75–80.

[3] M. E. Al-Mualla, C. N. Canagarajah, D. R. Bull, *Video Coding for Mobile Communications: Efficiency, Complexity and Resilience, Signal Processing and Its Applications*. Address: Academic Press, San Diego, CA, 2002.

[4] Wang, J. Ostermann, Y. Zhang, *Video Processing and Communications*. Address: Prentice Hall, 2001.

[5] J. R. Jain, A. K. Jain, "Displacement measurement and its application in interframe image coding," *IEEE Trans. Commun*, vol. 29 (12), pp. 1799–1808, Dec. 1981.

[6] L. Lurng-Kuo, E. Feig, "A block-based gradient descent search algorithm for block motion estimation in video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6 (4), pp. 419–422, 1996.

[7] T. Koga, K. Linuma, A. Hirano, Y. Iijima, T. Ishiguro, "Motion-compensated inter-frame coding for video conferencing," *Proc. NTC, New Orleans, LA*, 1981.

[8] R. Li, B. Zeng,M. L. Liou, "A new three-step search algorithm for block motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 4 (4), pp. 438–442, 1994.

[9] R. Li, B. Zeng,M. L. Liou, "A novel four-step search algorithm for fast block motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6 (3), pp. 313–317, 1996.

[10] J. Yeh, M. Khansari, M. Vetterli, "Motion compensation of motion vectors," *Proc. of IEEE International Conference on Image Processing*, vol. 1 , pp. 574–577, 1995.