RC24155 (W0701-093) January 19, 2007
Computer Science

IBM Research Report

Motion Estimation with Similarity Constraint and Its Application to Distributed Video Coding

Ligang Lu, Vadim Sheinin
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

Motion Estimation with Similarity Constraint and Its Application to Distributed Video Coding

Ligang Lu and Vadim Sheinin
IBM T. J. Watson Research Center
Yorktown Heights, NY 10598

ABSTRACT

In this paper we present a new motion estimation scheme that constraints the matching difference function with a similarity measure in its motion search process. We exploit the motion correlation among adjacent pixel blocks with similar statistics features and formulate this correlation as a similarity measure in terms of the motion similarity between the current pixel block and the neighboring blocks weighted by their corresponding statistical similarity. We then use this similarity measure as a constraint in the matching difference function to effectively trade off the differences in the pixel intensity values with the correlation in the adjacent motion vectors. Thus the motion estimation process becomes not only to minimize the differences in the pixel values but also to weight on the motion similarity in statistically similar neighbors. We applied this new motion estimation scheme to a distributed video coding system for side information generation at the Wyner-Ziv decoder and compared its performance to the scheme with no similarity constraint. The experimental results have shown that our motion estimation scheme with the similarity constraint has achieved significant gains in the fidelity of the generated side information and the decoded Wyner-Ziv frames over the scheme with no similarity constraint.

INTRODUCTION

Motion estimation has been widely used in predictive video coding and adopted by international standards[1,2] for generic motion picture coding. In a conventional, standard based video coding system, to encode a frame of video signal using motion compensated predictive coding, the encoder partitions the frame into blocks of pixels and performs block-based motion estimation and motion compensated prediction to exploit the temporal redundancy. In motion estimation, the encoder applies a search algorithm on a previously decoded adjacent frame (termed a reference frame) to find a reference block that minimizes the matching difference objective function measuring the distortion between the reference and the current block to be encoded. This reference block is the motion compensated prediction of the current block and the difference between the current block and the reference block is the motion compensated prediction error or the residue block. The encoder then applies transform coding on the residue block to make use of the spatial redundancy. For instance, the residue block is transformed into Discrete-Cosine-Transform (DCT) coefficients and the coefficients are then quantized and compressed further by entropy coding. Finally, the encoder formats the compressed coefficients, together with the coded corresponding motion vectors, into a bitstream according to the syntax and semantics defined by the standard for transmission or storage. Clearly, the motion estimation is a very important step and has a direct impact on the encoding performance; the better ability of the motion estimator has to find the true motion vectors, the higher performance the encoder can achieve. On the other hand, the ability of a motion estimator to find the true motion vectors depends on its search algorithm and search criterion. Over the years, researchers have investigated various motion estimation approaches [3, 4] and found that the block matching algorithm is more robust and has over all better performance and complexity ratio than other algorithms. If the search method and the block size are fixed (for example, if the full or exhaustive search method and 16x16 block) are used, then the motion estimation performance will be the function of the matching difference criterion defined according to some distance measure. The most commonly used matching criteria are defined by the sum of squared difference (SSD) and the sum of absolute difference (SAD) metrics. Minimizing SSD or SAD between the pixels in motion estimation is of mathematical simplicity and easy to implement. However, when there are changes in pixel intensity and noises, minimizing SSD or SAD often lead to false motion vectors. Furthermore, video often contains non-translational motions and moving objects are not necessary rigid due to camera zooming and viewing angle or position change. All these may affect the ability of the motion estimator to find the correct motion vectors. In this paper, we will present a new motion estimation scheme that employs a similarity constrained new search criterion to improve the motion estimation performance.

This paper is organized as follows. In the next section we will describe the new motion estimation scheme and in particular will formulate the similarity measure and incorporate it in the objective matching difference function as a constraint. Then in Section 3, we will apply our new motion estimation scheme in distributed video coding for side information generation at Wyner-Ziv decoder. In Section 4, we will evaluate the performance of our new motion estimation scheme and compare it with the performance of the motion scheme without the similarity constraint.

2. MOTION ESTIMATION WITH SIMILARITY CONSTRAINT

We propose a new motion estimation scheme to address the aforementioned drawbacks by exploiting the motion correlation existing in similar adjacent pixels. This is achieved by formulating this correlation in the form of a similarity measure in terms of the motion similarities among

statistically similar neighboring blocks and incorporating the similarity measure into the objective matching difference function as a constraint in the minimization process.

2.1 Motion Estimation with Conventional Minimization Criterion

Denote $B_{i_{1i}j_1}$ an *NxM* block of pixels in the current frame P_N and $B'_{i_2 j_2}$ an *NxM* block of pixels in the reference frame, where $i_k, j_k, k = 1,2$ are the indexes of the column and the row, respectively. We want to estimate the motion vector for $B_{i_1 j_1}$ by searching the reference frame P'_{N-1} to find the best matching block $B'_{i_2 j_2}$ that minimizes a difference function in a search window of appropriate size

$$J = \sum_{p_{xy} \in B_{i_1 j_1}, p'_{x'y'} \in B'_{i_2 j_2}} | p_{xy} - p'_{x'y'} |^n \qquad (1)$$

where p_{xy} and $p'_{x'y'}$ are intensity values of pixels at point (x, y) in P_N and point (x', y') in P'_{N-1}, respectively. In (1), when n=2 in (1), J is the well known SSD metric and when n=1, J is the widely used SAD metric. Using the difference function in (1) as the matching criterion, the motion estimator searches only to minimize the difference in the intensity value of the corresponding pixels. However such minimization criterion is sensitive to noises and light condition changes and may lead to false motion vectors.

2.1 Motion Estimation with Similarity Constrained Minimization Criterion

Motion vectors of pixels that belong to the same object will have high correlation. Judiciously exploiting such correlations among the neighboring motion vectors may effectively help to prevent or reduce the false motion vectors and outliers due to slow pixel intensity change and various types of noises (i.e., source acquisition noise, transmission noise, and coding noise). Our approach to make use of the correlations of the neighboring motion vectors is to measure the similarity between the candidate motion vector and the existing neighboring motion vectors and incorporate this similarity measure as a constraint in the search for the reference block that minimizes the matching difference criterion. However, neighboring blocks belong to different objects or background may likely have different motions; we need to distinguish the neighboring blocks and only utilize the motion correlations in those neighboring blocks that are similar to the current block $B_{i_{1i}j_1}$ and apply the motion similarity constraint accordingly in the minimization process.

Recently video similarity measurements have been studied and new similarity measures based on local statistics have been developed and successfully applied for video quality assessment [5, 6]. In has been shown in [5, 6] that local statistic features can provide a more noise robust and perceptually better measures on the similarity of two blocks of pixels than the conventional mean squared error based distance measure. Examples of these local statistic features include the mean, the variance, and the covariance between two blocks of pixels. These statistic features, and also others, can be incorporated into a similarity measure. Specifically, let *P* and *Q* are two blocks of pixels with a size of n × m, the pixel values in block *P* are denoted as P_{ij}, and the pixel values in block *Q* can be denoted as Q_{ij}, wherein i = 1, 2, …, n, and j = 1, 2, …, m. The sample mean, the sample variance of *P* and *Q*, and the covariance of *P* and *Q* are defined, respectively, as

$$\mu_P = \frac{1}{nm}\sum_{j=1}^{n}\sum_{i=1}^{m} P_{ij}, \quad \mu_Q = \frac{1}{nm}\sum_{j=1}^{n}\sum_{i=1}^{m} Q_{ij};$$

$$\sigma_P^2 = \frac{1}{mn-1}\sum_{j=1}^{n}\sum_{i=1}^{m} (P_{ij} - \mu_P)^2,$$

$$\sigma_Q^2 = \frac{1}{mn-1}\sum_{j=1}^{n}\sum_{i=1}^{m} (Q_{ij} - \mu_Q)^2;$$

$$\sigma_{PQ} = \frac{1}{mn-1}\sum_{j=1}^{n}\sum_{i=1}^{m} (P_{ij} - \mu_P)(Q_{ij} - \mu_Q).$$

For our application, we made a modification to the similarity index (SI) between *P* and *Q* in [5] as

$$SI(B_P, B_Q) = \alpha + \left[4\mu_P\mu_Q\sigma_{PQ} - \left(\mu_P^2 + \mu_Q^2\right)\left(\sigma_P^2 + \sigma_Q^2\right)\right]^2$$

where α is a positive constant. Similarity index so defined has the property

$$SI(B_P, B_Q) \geq \alpha,$$

with the equality if and only if $B_P = B_Q$. Now we present a scheme to use SI weighted the motion similarity to constraint the minimization process in the motion estimation. Let N be the set containing all the available neighboring blocks of the current block $B_{i_{1i}j_1}$. Let $mv_{i_1 j_1}$ be the motion vector being estimated for $B_{i_{1i}j_1}$ and let mv_{nm} be the motion vector associated with the neighboring block $B_{nm} \in N$, we define the new matching distance function for the motion estimation minimization criterion as

$$J = \sum_{p_{xy} \in B_{i_1 j_1}, p'_{x'y'} \in B'_{i_2 j_2}} | p_{xy} - p'_{x'y'} |^n + \lambda \sum_{B_{nm} \in N} SI(B_{i_1 j_1}, B_{nm}) \, | mv_{i_1 j_1} - mv_{nm} |^n . \qquad (2)$$

Where λ is a weighting factor reflecting the importance of the motion similarity constraint. Using the difference function defined in (2), we can reduce the effects of the pixel intensity changes caused by noises and light changes in the motion estimation. This new matching criterion will tend to smooth the motion vectors of the similar adjacent blocks and penalize the outliers with block similarity weighted

difference in motion vectors. In the next section, we will apply this new motion estimation scheme in distributed video coding wherein finding the correct motion vectors is more important than obtaining the minimum pixel difference in motion estimation for the side information generation in Wyner-Ziv decoding.

3. APPLICATION TO DISTRIBUTED VIDEO CODING

Recently significant research efforts have been devoted to develop practical distributed video coding (DVC) systems for emerging applications, such as distributed video surveillance system, mobile visual communication, etc. DVC encoders have very restricted computing power and need to adopt low complexity encoding algorithms while the decoders at a central location can have the resource for sophisticated signal processing operations. Current available video coding standards, such as MPEG-x and H.26x, are developed for traditional complex encoder-simple decoder system deployments and therefore are not suitable for DVC applications. The work by Slepian-Wolf and Wyner-Ziv have laid the theoretical ground for a low complexity encoding and high complexity decoding system, such as the DVC systems, to approach the same rate-distortion performance as the traditional coding systems. Slepian and Wolf showed for lossless coding case that given two correlated data sources X and Y, even if the side information Y is only available at the decoder, the best achievable rate to compress X is still the conditional entropy H(X/Y), the same rate as Y is also available at the encoder. Wyner and Ziv extended this result to the lossy compression case for Gaussian sources. Several recently published papers proposed distributed video communication systems based on Wyner-Ziv's theoretical work [7-9]. A key component for a DVC system to achieve efficient coding performance is the side information generation at the decoder. The higher the correlation between the source X and the side information Y, the better the coding performance can be achieved. Several papers have investigated the side information generation and presented progresses in this problem [10-12]. In this section we will apply our new motion estimation scheme in side information generation for the DVC system.

3.1 System Framework

Figure 1 depicts the general framework of our DVC system, the source X is split into two sub sources. The frames of one sub source are encoded by a Wyner-Ziv (W-Z) encoder. The frames of the other sub source are encoded by a traditional encoder, e.g., an H.264 encoder, and transmitted to the decoder and these frames are decoded by the traditional decoder and serve as the reference frames in side information generation. Then the Wyner-Ziv frame X' is decoded by exploiting the correlation between X and Y.

3.2 Side Information Generation

We apply our new motion estimation scheme described in the previous section for side information generation in our DVC system. Figure 2 provides a diagram description of the side generator. The side information generator consists of a similarity estimator, a motion estimator, and a motion compensated interpolator. The similarity estimator computes the similarity index (SI) for each of the existing neighboring blocks of the current block. The SI indexes are used to weight

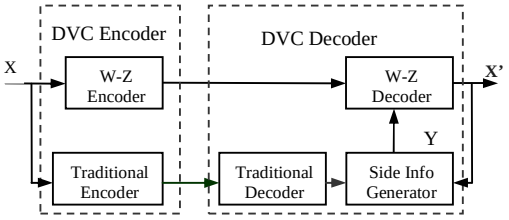

Fig. 1 Diagram of DVC System

the motion similarity constraint in the motion estimation algorithm. To generate the side information Y_N for decoding the current W-Z frame X_N, we use decoded frames for motion compensated interpolation. Note that the decoded frame can either be a decoded reference frame or a decoded W-Z frame. Motion estimation is first performed between X'_{N-1} and X'_{N+1} by minimizing the similarity constrained difference function in (2). Then the side information Y_N is generated by motion compensated interpolation. Empty pixel positions are filled using the method we reported in our previous paper [10]. Y_N is fed to the Wyner-Ziv decoder and used to decode X_N.

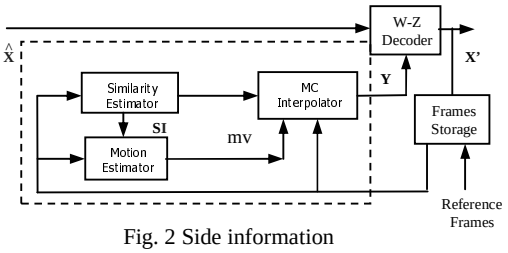

Fig. 2 Side information

4. SIMULATION RESULTS

In this section we evaluate the performance of our new motion estimation scheme presented in Section 2. Since the well known full search block matching motion estimation generally has the best performance among all block based motion estimation schemes, we chose the full search block matching motion estimation scheme with the difference function in (1) with n=1 as our benchmark for comparison. Our similarity constrained motion estimation scheme is implemented using the difference function in (2) with n=1; in the experiment, the weighting factor λ for the similarity constraint is 28.0. All motion estimations for both schemes are done in quarter-pel resolution. We applied both motion estimation schemes to the DVC system described in the last section to estimate the motion vectors between the decoded references frames. The motion vectors are then used in motion compensated interpolation to generate the side information for decoding the current Wyner-Ziv frame. At the DVC encoder, all the odd number frames were encoded by a simplified H.264 P-frame encoder at the constant quant step size of 30 and send to the DVC decoder. All the even number frames were Wyner-Ziv coded. The decoded H.264 frames are used as the reference frames to generate the side information for decoding the Wyner-Ziv frames.

To show the advantage of our motion estimation scheme with the similarity constraint, we compared the quality of the side information generated by our scheme and by the benchmark scheme and the quality of the decoded Wyner-Ziv frames using the two schemes. Figure 3 and Figure 4 are the peak-signal-to-noise ratio (PSNR) comparisons on the standard CIF format video test sequences of Mobile Calendar and Coastguard. All PSNR are calculated against the original sources. The plots in Figure 3 have shown that, in Mobile Calendar case, our new motion estimation scheme has outperformed the bench mark scheme with a gain on average of 1.02 dB (30.68 dB vs. 29.66 dB) and up to 2.23 dB in side information generation. It also achieved a gain on average of 0.7 dB (35.73 dB vs. 35.03 dB) and up to 1.89 dB on the decoded Wyner-Ziv frames. Figure 4 presents the results on Coastguard. Again, the plots have shown that our new scheme outperformed the bench mark scheme in the side information generation with a gain on average of 0.9 dB (33.07 dB vs. 32.17 dB) and up to 1.27 dB and a gain of 0.23 dB on average and up to 0.33 dB on the decoded Wyner-Ziv frames. Here we should point out that, comparing to the gains in side information generation, the gains in decoded Wyner-Ziv frames are relatively smaller. This is due to model mismatch and other imperfectness in the Wyner-Ziv coding.

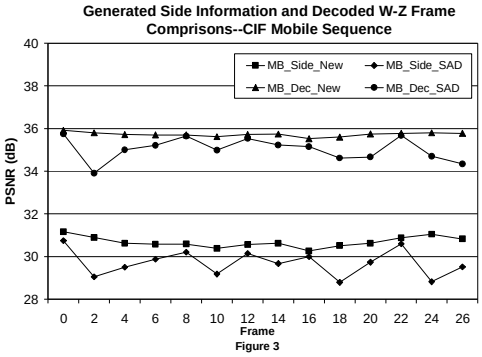

Figure 3

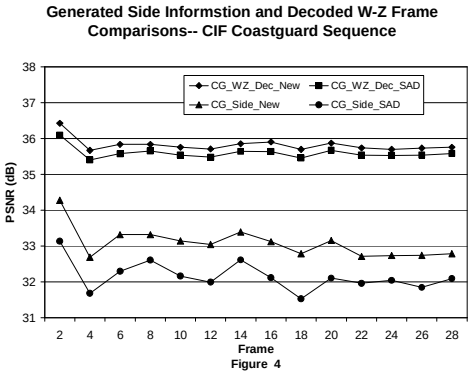

Figure 4

5. Summary and Conclusions

In this paper we have presented a new motion estimation scheme. This scheme can exploit the motion correlation that exists in neighboring similar blocks. The similarity measure is formulated in terms of the motion similarity weighted by a local statistical similarity. The similarity measure is used as a constraint in the objective difference function to reduce the effects of noises and pixel intensity changes and to trade off the minimization of the difference in pixel values with the motion smoothness in the motion estimation. We also have applied this new motion estimation scheme in a DVC system to generate side information for Wyner-Ziv decoder and compared its performance with a bench mark scheme. The results have shown that our motion estimation scheme has achieved significant gains in the fidelities of both the side information and decoded Wyner-Ziv frames.

REFERENCES

1. ISO/IEC 13818-2 Generic Coding of Motion Pictures and Associated Audio: Video, 1995.

2. ISO/IEC 14496-2 Generic Coding of Audio-Visual Objects-Part 2 Visual, 2000.

3. F. Dufaus and F. Moscheni, "Motion estimation techniques for digital TV: A review and a new contribution," *Proc. IEEE*, vol. 83, pp.858-876, Jun. 1995.

4. A. M. Tekalp, *Digital Video Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1995, pp.72-129.

5. L. Lu, Z. Wang, A. Bovik, "Full Reference Video Quality Assessment Considering Structural Distortion and No-Reference Quality Evaluation of MPEG Video", *Proceedings of IEEE ICME*, Lausanne, Switzerland, Aug. 22~25, 2002.

6. Z. Wang, L. Lu, and A. C. Bovik, *Video quality assessment based on structural distortion measurement*, Elsevier Signal Processing: Image Communication., vol. 19, pp. 121-132, 2004.

7. R. Puri and K. Ramchandran, "PRISM: A New Robus Video Coding Architecture based on Distributed Compression Principles", in *Allerton Conference on Comm, Control, and Computing*, vol.1, pp. 240-244, Urbana-Champaign, IL, Oct. 2002.

8. A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv coding of motion video," *Proc. of the 36th Asilomar Conf. on Signals, Systems, and Computers*, Pacif Grove, CA, Oct. 2002.

9. B. Girod, A. Aaron, S. Rane, and D. R. Monedero, "Distributed Video Coding," *Proc. of the IEEE*, vol. 93, pp. 71-93, Jan. 2005.

10. L. Lu and V. Sheinin, "Side Information Generation for Low Complexity Video Coding Systems Based on Wyner-Ziv Theorem," *IEEE Intl. Symposium. On Broadband Multimedia Systems* , Las Vegas, NV, April 2006.

11. S. Klomp, Y. Vatis, and J. Ostermann, "Side Information Interpolation with Sub-Pel Motion Compensation for Wyner-Ziv Decoder," *Intl. Conf. on Signal Processing and Multimedia Applications*, Setubal, Portugal, Aug. 7-10, 2006.

12. L. Natario, C. Brites, J. Ascenso, and F. Pereira, "Extrapolating side information for low-delay pixel-domain distributed video coding," *Intl. Workshop on Very Low Video Coding*, Sardenha, Italy, Sept, 2005.

13. L. Zhen and E. J. Delp, "Wyner-Ziv video side estimator: conventional motion search method revisited." *Proc. of ICIP*, pp. 825-828, Genova, Italy, Sept. 2005.