BUILDING A PSEUDO OBJECT-ORIENTED VERY LOW BIT-RATE VIDEO CODING SYSTEM FROM A MODIFIED OPTICAL FLOW MOTION ESTIMATION ALGORITHM

Chung-Wei Ku, You-Ming Chiu, Liang-Gee Chen, and Yung-Pin Lee

DSP/IC Design Lab., Department of Electrical Engineering,
National Taiwan University, Taipei, Taiwan, R.O.C.
email: william@video.ee.ntu.edu.tw

ABSTRACT

In this paper, a modified optical flow algorithm (MOFA) is proposed for the development of a very low bit-rate video coding system. Another edge preserving constraint and pyramid approach are suggested to generate an accurate motion field and reduce the possibility of trapped in local minimum respectively. Post-processing schemes are also designed to eliminated the problems for special regions. Compared with other motion estimation algorithms, the proposed method gives more exact estimation in terms of PSNR and subjective view. The arbitrarily shaped transform of motion field is then selected to further remove spatial redundancy. According to some primary simulation results, most of the test sequences are compressed into 16 Kbps or lower with excellent picture quality.

1. INTRODUCTION

Although H.261 has defined a standard for videophone or videoconferencing with $p\times 64$ Kbps, lower bit-rate is expected to utilize current telephone network. Because the channel bandwidth is small, a very high compression ratio must be achieved. According to the standard of modern modem (V.34), transmission rate is defined as 28.8 Kbps. As early as May 1991 the MPEG raised the issue of audio-visual standard targeted at the bit-rate of 4.8-64 Kbps, with the motivation being the limited channel bandwidth and limited storage capacity. These efforts was approved in July 1993 with the MPEG-4 nickname and the title "Very Low Bit-Rate Coding of Moving Pictures and Associated Audio". Recently, ITU-T announced a draft about "Video Coding for Narrow Telecommunication Channels at below 64 Kbit/s" or H.263. It is still based on a block-wise coding algorithm basically. In fact, it is a widespread belief that only substantial innovations in coding algorithm will produce a lasting standard which can satisfy users.

In order to compress video signal efficiently, motion compensation (MC) is widely adopted in many standards where fix-sized block matching algorithm is applied. However, for very low bit-rate video coding visible block effect will degrade the performance. To tackle this problem, many quite different approaches are proposed; such as model-based coding [1] and analysis-synthesis coding algorithms [2]. Basically speaking, object-oriented coding is block-effect free and its visual performance is good enough. In addition, it is an efficient approach since head and shoulders are the major parts on screen for conversation applications. It also provides more meaningful information compared with traditional waveform coding scheme. In this paper, we propose a modified optical flow algorithm (MOFA) for motion estimation. Compared with other motion estimation algorithms, MOFA gives accurate motion field and its performance is efficient for video coding or image understanding. Because the motion field generated by MOFA is homogeneous, the "pseudo objects" are extracted and segmented easily. These objects are applied with arbitrarily shaped transform (AST) to achieve data compression further.

2. MODIFIED OPTICAL FLOW ALGORITHM

Although optical flow algorithm can generate a reasonable motion field, some drawbacks still exist. In order to code the motion vectors more efficiently, we would like to generate a motion field with less spatial variation but with good quality at the same time. Instead of using extra estimation filtering scheme, we propose a modified cost function which is shown as follows:

$$\varepsilon = \int\int \quad ((E_x u + E_y v - E_t)^2 + \alpha(u_x^2 + u_y^2 + v_x^2 + v_y^2) \\ +\beta(u_{xy}^2 + v_{xy}^2))dxdy.$$

0-7803-3192-3/96 $5.00©1996 IEEE

The additional third term indicates the penalty on convoluted edges. In other word, the cost function favors motion fields with straight edges. To minimize the cost function, the following relation must be satisfied:

$$\frac{\partial \varepsilon}{\partial u} - \frac{\partial}{\partial x}\frac{\partial \varepsilon}{\partial u_x} - \frac{\partial}{\partial y}\frac{\partial \varepsilon}{\partial u_y} + \frac{\partial^2}{\partial x \partial y}\frac{\partial \varepsilon}{\partial u_{xy}} = 0.$$

By similar derivation, we have

$$4(\alpha\overline{u} + \beta\hat{u} - (\alpha+\beta)u) = \frac{1}{\lambda}(E_x u + E_y v - E_t)E_x,$$

$$4(\alpha\overline{v} + \beta\hat{v} - (\alpha+\beta)v) = \frac{1}{\lambda}(E_x u + E_y v - E_t)E_y.$$

where $\hat{u}$, $\hat{v}$ indicates local averages. Let

$$\alpha\overline{u} + \beta\hat{u} = (\alpha+\beta)u^*, \qquad \alpha\overline{v} + \beta\hat{v} = (\alpha+\beta)v^*,$$

then

$$u_{i,j} = u^* - \tfrac{E_x u^* + E_y v^* - E_t}{4(\alpha+\beta)+E_x^2+E_y^2}E_x,$$
$$v_{i,j} = v^* - \tfrac{E_x u^* + E_y v^* - E_t}{4(\alpha+\beta)+E_x^2+E_y^2}E_y. \qquad (1)$$

Another problem is that the answers may be trapped in local minimum due to gradient descent approach. To reduce the probability of this situation, a pyramid approach is suggested. The higher level of the pyramid is composed of the subsampled pixels at the lower level. Motion estimation is executed from the top level to the bottom level. The initial guess on each level is the results of its adjacent higher level. The scale of adjustment at higher level is larger than that of lower ones. At higher level, the local details are filtered out so the possibility of trapped in local minimum is reduced; at lower level, the motion vectors are adjusted locally to make more precise estimation. Therefore, at lower level we do not need to execute too many iterations. Generally most local minima are removed while the global minimum is reserved.

For gradient descent approaches, large prediction errors occur on the edges due to the large gradients; the motion vectors in edge regions must be processed in an alternative way. If the compensated error is too large, the motion vector of this pixel will be replaced by one motion vector of its neighbors. Figure 1 illustrates an example: suppose v_0 indicates the motion vector of the current pixel, and v_1, v_2, v_3, v_4 are the motion vectors of its 4 neighbors. If the residual error of v_0 is larger than a predefined threshold, the 4 residues, denoted as r_1, r_2, r_3, r_4 which are the corresponding residual errors of v_1, v_2, v_3, v_4 for the current pixel, are compared one another. In case r_1 is the minimum, the motion vector for the current pixel will be replaced by v_1. Since the original pixel with motion vector v_1 is not in the edge region, in general its estimated vector should be more accurate than the others. It is reasonable that many motion vectors in edge regions will be substituted for the better vectors which are found in the homogeneous regions. We find the substitution scheme alleviates most of the performance degradations due to edges. On the other hand, in the flat stationary regions such as background, there may be some non-zero motion vectors close to the moving edges because the motion vectors of the moving part will "propagate" to their neighboring stationary pixels due to the smoothness constraint. Therefore, these non-zero parts must be filtered out to generate a more homogeneous and practical motion field. We can apply a large Gaussian window convolved with the residual error; if the result is less then the window convolved with frame difference (error of non-motion), then set the motion vector of the central pixel to be zero. In other word, given a Gaussian window G and residual error R, if $G \cdot R < G \cdot R_0$, where R_0 is frame difference, then the motion vector of the central pixel in G will be forced to be zero because the region around the pixel is stationary.

All the above we call it a *modified optical flow algorithm* (MOFA). To prevent the drawbacks of optical flow algorithms, the modified optical flow algorithm utilizes extra edge preserving constraint and pyramid approach to generate an edge-preserving motion field and alleviate the risk of local minimum. Furthermore, the post processing of edges and stationary regions guarantees the motion field to be more realistic for either video coding or object extraction, as later simulation results show.

3. SIMULATION RESULTS OF THE MOTION ESTIMATION ALGORITHMS

To understand how the mentioned motion estimation algorithms perform, a 176 × 144 with 10 frames/sec sequence "Miss America" is selected as the test sequence. Four kinds of motion estimation algorithms, block matching algorithm, pel-recursive algorithm, optical flow algorithm, and modified optical flow algorithm, are applied on the successive frames to generate the motion fields, respectively. The block matching algorithm is full-search of 16 × 16 blocks with searching range -8 to +7. The motion vectors within a block are all the same for block matching algorithm obviously. For pel-recursive algorithm, the distribution of motion vectors is too random to be encoded efficiently. In addition, most of the motion vectors around edge region are trapped in local minimum. Original optical flow algorithm with a smoothness constraint can generate a more reasonable motion field, but some vectors are still trapped in local minimum. Besides, the evaluated motion vectors seems to be not large enough for such a fast movement. The proposed modified optical flow algorithm with an edge-preserving constraint and pyramid-search approach generates the best motion field; the movement of object can be observed and the motion field can be encoded the most efficiently, as an example in Figure 2 shows. Table 1 lists a summary of these algorithms. For all the test sequence, MOFA gives the best estimation and the generated motion field is efficient for both coding and understanding. In sequence "Elsa" and "Caisy", which are also "head and shoulders" sequences with small motion, the movement is not very sharp so the improvement of MOFA is only about 1dB compared with OFA. However, for sequence with large motion such as "Miss America" and "Jian", the improvement of performance is significant.

4. PROPOSED PSEUDO OBJECT-ORIENTED VIDEO CODING SYSTEM

The first frame is encoded in intra mode; this part is similar to H.261 and not displayed in the figure. For inter frame coding, motion estimation between the current and previous frames is applied according to the modified optical flow algorithm. As previously mentioned, the proposed MOFA will generate a dense motion field. Because of the homogeneous property of the motion field , we can segment the motion field into several "pseudo objects" easily; in fact, there is only one major "pseudo object" in our applications and we call this object as *Motion Compensated Object* (MCO). Of course some areas on screen could not be encoded exactly by motion compensation. These areas are named as *Motion Off Objects* (MFO's). The coding scheme of MFO is still under development. All the motion vectors generated by MOFA will be segmented into regions which are further compressed by arbitrarily shaped transform (AST) [6]. AST is very similar to DCT except that the transformed region is not restricted to rectangles. The price paid for is the information about the shape should be transmitted because the transformation kernels depend on the shape of the region. The operation of AST includes: find the circumscribed rectangle of the encoded region, orthogonalize the transformation kernels according to the shape of the region, and apply AST to the x and y parts of the motion field respectively. Finally the parameters about the shape and two groups of transformed coefficients are sent to the variable length coder for lossless compression. A feedback loop reconstructs the picture for the synchronization with decoder. Getting together with these coded elements, a control buffer arranges the data stream and set the priority flag for each object. In decoder, just the reversed operations of the above are applied. Because the proposed coding method is based on applying AST to the motion field of the "pseudo object" which is extracted by MOFA, the whole system is called "pseudo object-oriented video coding system" [5]. However, it is different from previous object-oriented approaches. Currently the whole system is written in C language and built in X-Window environment. We develop all the interfaces and make the system user-friendly in an interactive style. According to the current simulation results of our system, we can easily compress the test sequences into 16 Kbps or even lower. At the same time, the quality is still guaranteed. We are trying to improve the operations in AST and the process about MFO now. In fact, all primary results have shown that our proposed methods will be very useful for very low bit-rate video coding applications. Table 2 gives the average performance of the proposed system for several test sequences.

5. CONCLUSION

In this paper, we propose a modified optical flow based motion estimation algorithm. To eliminate the block effect in block matching approaches, a pixel based motion estimation algorithm is suggested. To avoid the weakness of all the other pixel based algorithms, we design several schemes in our modified optical flow algorithm. The suggested edge preserving term in cost function guarantees a more realistic motion field. The pyramid approach reduces the possibility of trapped in local minimum. All these strategies make the estimation faster and more accurate even for large movements. We also designed a post-processing method which improves the accuracy of motion vectors in both edge and smooth regions. Compared with other motion estimation algorithms, its performance is the best for both subjective view and PSNR. Besides, the generated motion field is more practical for video coding or image understanding.

In the proposed system, MOFA is chosen as the motion estimation algorithm to remove the temporal redundancy in video sequence. The generated motion field is segmented into regions and applied with arbitrarily shaped transform to remove the spatial redundancy in motion field. For conversation application, usually there is only single object on screen and most of the contents on screen are well compensated by the above method; this "pseudo object" is called motion compensated object (MCO). The process of MFO should be further improved by using some VQ methods. Generally speaking, our system is a "pseudo object-oriented" approach for very low bit-rate video coding, but different from the object-oriented methods. Since the patterns appear on screen is less restricted in our system, we believe the proposed system is more suitable and practical for videophone or videoconferencing.

In the future, we will combine this motion estimation algorithm with the AST technique which is modified and improved currently to achieve more data compression. To optimize the system, it is interesting to investigate the possibility of optimizing MOFA and AST all together rather than individually. In addition, for some detail operations such as the actions of eyes and mouth, several fine compensation methods can also be appended to advance the acceptance of picture. The short term goal of our group is to build a prototype of very low bit-rate video coding system. For a long term point of view, we will try to build a multimedia email system, or a multimedia telephone system.

6. REFERENCES

[1] K. Aizawa, H. Harashima and T. Saito, "Model-based analysis-synthesis image coding (MBASIC) system for a person's face", *Signal Processing: Image Communication*, vol. 1, pp. 139-152, 1989.

[2] H. Shiller and M. Hötter, "Investigations on color coding in an object-oriented analysis-synthesis coder", *Signal Processing: Image Communication*, vol. 5, pp. 319-326, 1993.

[3] D.R. Walker and K.R. Rao, "Improved pel-recursive motion compensation", *IEEE Trans. Communication*, col. COM-32, pp. 1128-1134, OCT. 1984.

[4] B.K.P. Horn and B.G. Schunck, "Determining optical flow", *Artificial Intelligence*, Vol. 17, pp. 185-203, 1981.

[5] C.-W. Ku, L.-G. Chen, and Y.-M. Chiu, "A Very Low Bit-Rate Video Coding System based on Optical Flow and Region Segmentation Algorithms", *Proceeding of the SPIE Visual Communication and Image Processing, Taipei*, vol. 3, pp. 1318-1327, May 1995.

[6] M. Gilge, "Coding of arbitrarily shaped image segments based on a generalized orthogonal transform", *Signal Processing: Image Communication*, vol 1, pp. 153-180, 1989.

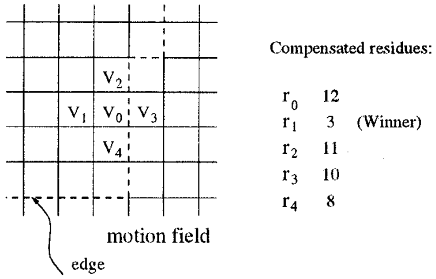

Figure 1: Example of vector substitution on the edge (v_0 is substituted for v_1).

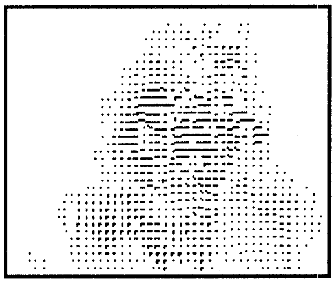

Figure 2: The motion field generated by MOFA.

Table 1: The performance of several algorithms.

Sequence name	BMA	PRA	OFA	MOFA
Miss America	34.525	37.353	38.471	41.175
Elsa	36.666	38.751	41.149	42.662
Jian	32.705	36.919	38.693	40.856
Caisy	37.779	39.488	43.019	43.624

Table 2: Primary results of the proposed system.

Sequence name	Sequence length	PSNR	Bit-Rate
Miss America	47	34 dB	16 Kbps
Elsa	51	34 dB	9 Kbps
Caisy	32	35 dB	9 Kbps
CMJ	530	34 dB	7 Kbps