

OBJECT-ORIENTED CODING USING SUCCESSIVE MOTION FIELD SEGMENTATION AND ESTIMATION

Dam LeQuang, André Zaccarin and Sylvain Caron

Dépt. de génie informatique et génie électrique
Université Laval
Ste-Foy, Québec, Canada, G1K 7P4
e-mail: lequang@gel.ulaval.ca zaccarin@gel.ulaval.ca
tel: (418) 656-2130, fax: (418) 656-3159

Abstract - Block-based motion compensation assumes that all pixels within a block have the same translational motion. That hypothesis, however, results in inaccurate compensation of moving objects' boundaries. Object-oriented video compression algorithms typically segment each image in regions of uniform motion and estimates the motion of these regions to generate more accurate motion compensated images. In this paper, we present a two-stage algorithm for motion field segmentation and estimation in an object-oriented coder. In the algorithm's first stage, a standard block-matching algorithm and a maximum a posteriori probability estimate are used to compute a translational motion field and its segmentation. This segmentation is then utilized in the second stage to estimate the parameters of complex motion models. Parameters of complex motion models are only estimated in the algorithm's second stage which reduces the computational complexity of the proposed algorithm. Simulation results show that the proposed algorithm significantly reduces the bit rate needed to encode video sequences when compared to standard block-based algorithms.

1. INTRODUCTION

Standard compression algorithms for video sequences, like H.261 and MPEG, use block-based motion compensated prediction to remove the temporal redundancy in video sequences. Block-based motion compensation is based on the hypothesis that blocks of pixels move with constant translational motion so that only one motion vector per block has to be transmitted. Each frame of a sequence is divided into blocks, typically of 16 by 16 pixels, and for each block a motion vector is estimated at the encoder and transmitted to the decoder so that a prediction of the present frame can be constructed from the previously decoded frame. Assuming the motion field to be constant over blocks, however, results in inaccurate compensation of moving objects' boundaries. Blocks located at these boundaries contain regions moving in at least two different directions and a single motion vector leads to incorrect compensation for at least one of these regions. A high bit rate is therefore needed to encode the prediction error so that blocking artifacts do not appear in the decoded sequence.

Several object-oriented video compression algorithms were proposed to compensate for the weaknesses of block-based approaches, e.g. [1-3]. Instead of estimating the motion of pixel

blocks, these algorithms segment each image in regions of uniform motion and estimate the motion of these regions. Although the motion field segmentation has to be encoded along with the parameters describing the motion of each region, that information allows the generation of better motion compensated predicted images. Consequently, a smaller bit rate is required to encode the prediction error, and the overall bit rate needed to encode a sequence with an object-oriented coder is usually less than that required by a block-based coder.

1.1. Previous work

Object-oriented coding requires the estimation of a motion field and the segmentation of the same motion field into regions of uniform motion. Several approaches have been proposed to solve the problem of motion field segmentation and motion field estimation. Musmann *et al.* [1] proposed an iterative approach in which an image is first divided into unchanged and changed areas. Parameters describing the motion of each changed area are computed. Within these areas, regions of connected pixels whose motion is not well described by the estimated motion parameters are detected, and their motion parameters are estimated. Wang and Adelson [2] proposed an approach for which motion parameters are first estimated and then used to compute a segmentation of the image to encode. To do so, an optical flow motion field is computed for the whole image. Linear regression is then used to estimate affine motion parameters on windows of 20 by 20 pixels. Valid motion parameters are identified by clustering in the parameter space and the image is then segmented into regions using these motion parameters. Chang *et al.* [3] use a Gibbs distribution to model the joint distribution of the motion field and its segmentation. A maximum a posteriori (MAP) estimate of the motion field and its segmentation, given the past and present images, is computed. Dense motion field estimation, parametrization of that motion field with complex motion models and motion field segmentation are computed simultaneously in an iterative algorithm.

In this paper, we present a new approach for motion field estimation and segmentation in an object-oriented image coder. In the proposed system, motion field segmentation is first performed using a simple translational motion model. That segmentation is then used in the second stage of the algorithm to estimate the parameters of more complex motion models, such as affine models. By delaying the use of complex motion models to the algorithm's second stage, parameter estimation is only performed once, therefore reducing the algorithm's complexity.

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada and by the Fonds FCAR (Prov. Québec).

2. PROPOSED OBJECT-ORIENTED ALGORITHM

As mentioned in the previous section, we propose a two-stage algorithm to solve the motion field estimation and segmentation problem. In the first stage of the algorithm, we estimate the segmentation of a motion field computed with a standard block-matching algorithm. In the second stage, affine motion parameters are computed for each segmented region from a motion field computed by optical flow. The block diagram of the proposed object-oriented coding algorithm is shown in Figure 1.

2.1. Segmentation of the motion field from block motion vectors

Previous work [4,5] shows that a motion field segmentation can be used to generate motion compensated images of good quality, even when the motion field is computed with a standard block-matching algorithm. In [4,5], each image is divided into blocks of 16 by 16 pixels and the motion field of each block is segmented into two regions, such that each region is well compensated by the motion vector of the block itself or a neighboring block. The motion field segmentation is modeled by a Gibbs distribution and a MAP estimate of that segmentation is computed. In the proposed approach, we use the same Gibbs distribution to model the segmentation, but we now compute the MAP estimate of the segmentation for the whole image, instead of computing a segmentation for each block of pixels.

Given the present image I^n , its previous image I^{n-1} and the motion field V_S that is initially calculated by a standard block matching algorithm, the motion field segmentation s is estimated by maximizing the conditional probability of the segmentation as follows

$$\hat{s} = \underset{s}{\operatorname{argmax}} \{P(S = s | I^n, I^{n-1}, V_S)\} \quad (1)$$

Using Bayes theorem, and after some simplifications, the segmentation estimate is given by

$$\hat{s} = \underset{s}{\operatorname{argmax}} \{P(I^n | S = s, I^{n-1}, V_S)P(S = s)\} \quad (2)$$

The first term is the conditional probability of the present image given the previous image, the motion field and the segmentation. Assuming that the motion field and the segmentation are correctly estimated, we can model the displaced frame difference (DFD) at each pixel p , given a motion vector v , as a Gaussian random variable with mean zero and variance σ^2

$$\text{DFD}^n(p, v) \sim N[0, \sigma^2] \quad (3)$$

The second term is the probability distribution of the motion field segmentation. In this paper, we use the MRF model previously used in [4,5]. In that model, an 8-pixel neighborhood is used. The prior distribution of the motion field segmentation is given by

$$P(S = s) = \frac{1}{Z} \exp \left[\sum_p -\beta \left(t_1(p) + \frac{t_2(p)}{\sqrt{2}} \right) \right] \quad (4)$$

where Z is a normalizing factor, β is a parameter used to adjust the relative strength of the smoothness constraint, $t_1(p)$ is the number of horizontal and vertical neighboring pixel pairs in the neighborhood of pixel p with different segmentation values, and

$t_2(p)$ is the number of diagonal neighboring pixel pairs with different segmentation values. The segmentation can therefore be estimated by finding the minimum of the following cost function

$$\hat{s} = \underset{s}{\operatorname{argmin}} \sum_p \left\{ \left[\text{DFD}^n(p, v) \right]^2 + \beta \left(t_1(p) + \frac{t_2(p)}{\sqrt{2}} \right) \right\} \quad (5)$$

We minimize that function by using a multi-resolution iterated conditional modes (ICM) algorithm [6] which significantly accelerates convergence while giving accurate segmentations.

Once a motion field segmentation is computed, the motion vector of each connected region is updated by matching the region of the present frame to the past frame, similarly to what is done in block matching. The search area is limited to ± 2 pixels around the motion vector of the region. It is not necessary to search in a larger area since the pixels in the same region are typically well compensated by the original block motion vector. Once new region motion vectors have been computed, it is possible to update the motion field segmentation. In our simulations, however, we found that this step was not necessary. Although the segmentation is computed using an iterative algorithm, a translational motion model is used at this stage of the algorithm to facilitate the computation of the motion compensated images. If a higher order motion model is used, such as an affine or quadratic model, the computational complexity of the segmentation increases significantly.

2.2. Computation of motion parameters from an optical flow motion field

After the algorithm's first stage, a segmentation of the image into regions of uniform motion has been estimated, and translational motion vectors describing the motion of each region have been computed. If that information is used to generate motion compensated images in a coding system, the algorithm's performance is comparable to that of a standard block-based coder because of the information overhead needed to encode the segmentation. To take advantage of that segmentation, the motion has to be modeled with higher order models, and sub-pixel accuracy is needed for the motion parameters.

We estimate the parameters of the complex motion models for the segmented regions as follows. First, a motion field for the whole image is computed by optical flow using the Horn and Schunck algorithm [7]. We iteratively compute the optical flow of the image by using the following equations

$$\begin{cases} v_x^{k+1} = \bar{v}_x^k - \frac{E_x \bar{v}_x^k + E_y \bar{v}_y^k + E_t}{\alpha^2 + E_x^2 + E_y^2} E_x \\ v_y^{k+1} = \bar{v}_y^k - \frac{E_x \bar{v}_x^k + E_y \bar{v}_y^k + E_t}{\alpha^2 + E_x^2 + E_y^2} E_y \end{cases} \quad (6)$$

where k is the iteration number, \bar{v}^k is neighborhood average of v . E_x , E_y , E_t denote spatio-temporal derivatives of image intensity E . The algorithm is initialized by assigning to each pixel the translational motion vector of its region that has been calculated in the algorithm's first stage. The optical flow is computed separately for each segmented region of the motion field. Therefore, no smoothness constraint is imposed across

regions' borders.

For each segmented region, we then use the motion vectors obtained by optical flow to compute a least-square estimate of the parameters of the desired motion model. In our work, we use an affine motion model (6 parameters). Simulations show that the parameters obtained with this approach model well the motion of each region. Estimating the parameters of a complex motion model is computationally intensive. As several algorithms that computes dense motion fields, the Horn and Schunck algorithm is iterative. In our coding system, however, the estimation of complex motion model parameters is performed only once, which reduces the overall computational complexity of the proposed algorithm.

The proposed approach tends to oversegment the motion field because the segmentation is performed with block motion vectors. This approach, however, allows us to estimate complex motion parameters only once. To compensate for the oversegmentation of the motion field, we have implemented an algorithm that merges neighboring regions whose motion parameters are similar. Based on rate-distortion theory, our merging criterion is such that we can reduce the number of regions while keeping the increase in prediction error as small as possible. The merging criterion is as follows

$$[D_{ij} - (a_i D_i + a_j D_j)] < \lambda [(R_i + R_j) - R_{ij}] \quad (7)$$

where λ is non-negative, and ij is the index of the region obtained by merging region i and j . For each region, D is the prediction error, and R is the total number of bits needed for encoding that region. Depending on the value of λ and the considered image content, the number of regions can typically be reduced by 30% - 40% without significantly increasing the prediction error.

Each image is therefore represented by 3 sets of parameters: (M) motion, (S) segmentation and (P) prediction error. The motion parameters are encoded with a precision of $1/32$. All zero parameters are marked and encoded with only 1 bit. We encode precisely the region contours of the motion field segmentation by using Freeman coding and a Markov model of order 1. The average number of bits required for encoding a pixel of the contour is typically 1.6 bits. The last set of parameters, the image prediction error, is divided in blocks of 8 by 8 pixels, transformed by DCT, quantized, run-length and Huffman encoded, as done in standard coding algorithms like JPEG and MPEG.

3. RESULTS AND CONCLUSION

Figure 2 shows the results of a simulation run on the first 30 frames of the *Table-tennis* sequence in Common Intermediate Format (CIF). Frames 0, 10 and 20 were intra coded. Simulations were also run on the *Football* and *Flower garden* sequences and similar results were obtained. In the figure, BMA stands for a standard block-matching algorithm for which motion vectors were computed with half-pixel accuracy. SEG corresponds to the first stage of the proposed algorithm, i.e., an algorithm for which each segmented region of the motion field is only compensated by the half-pixel accuracy motion vector computed by region matching. Finally, OOC is the proposed object-oriented coding algorithm.

Using BMA, SEG and OOC, the *Table-tennis* sequence was encoded so that the 3 decoded sequences had comparable mean squared error (MSE). At frame rate (30 frames/sec), the 3 decoded

sequences were also visually similar. Figure 2 gives the MSE between the original images and the decoded images, as well as the bit rate needed to encode the images. As that figure shows, the performance of the proposed algorithm, OOC, is significantly better than that of a standard block-based coder (BMA). On average, our object-oriented coder requires 20% - 35% fewer bits than the BMA algorithm. It is also interesting to note that the performance of the SEG algorithm is only comparable and sometimes worse than BMA. For SEG and OOC, an information overhead is needed to encode the motion field segmentation. In SEG, however, a translational motion model is used with half-pixel accuracy motion vectors. That precision is not sufficient to eliminate motion compensated prediction errors on the object contours. With OOC, better motion models and the accuracy of the motion parameters eliminate most of these errors, which in turn significantly reduce the bit rate. Table 1 gives the reduction in bit rate obtained on simulations run on 3 video sequences, when compared to BMA. Table 1 also gives the average number of segmented regions, as well as the percentage reduction in the number of regions obtained with the merging algorithm.

Table 1: General simulation results

Video sequence	Gain in bit rate (%)	Number of regions	Reduction in number of regions (%)
<i>Tennis</i>	36	25	40
<i>Football</i>	20	70	33
<i>Flower</i>	24	65	30

Figure 3 is a section of a motion compensated predicted image obtained with BMA and OOC. As these images show, the proposed algorithm generates motion compensated predicted images of very high quality. In the current implementation of the OOC algorithm, the prediction error is encoded using DCT with a fixed quantization table. We believe, however, that with a better approach to encode the prediction error, our algorithm could be used for coding applications at very low bit rates. This is also supported by the fact that for the simulation results shown in Figure 2, only 25% of the bit rate is used to encode the motion field segmentation and the motion parameters. The distribution of the bit rate among the 3 sets of parameters is given in Table 2 for 3 video sequences.

Table 2: Distribution of the bit rate

Video sequence	M (%)	S (%)	P (%)
<i>Tennis</i>	5.8	17.6	76.6
<i>Football</i>	6.7	21.2	72.1
<i>Flower</i>	5.3	19.2	75.5

The approach that we propose in this paper can be thought of as an algorithm that hierarchically estimates parameters of complex motion models in an object-oriented coder. Such an approach reduces the computational complexity of object-oriented coders, and our simulation results show that an algorithm with a two-stage hierarchy significantly reduces the bit rate needed to

encode a video sequence when compared to standard block-based algorithms.

4. REFERENCES

- [1] H. G. Musmann, M. Hötter and J. Ostermann, "Object-oriented analysis-synthesis coding of moving images", *Signal Processing: Image Communication*, vol. 1, pp. 117-132, 1989.
- [2] J. Y. A. Wang and E. H. Adelson, "Spatio-temporal segmentation of video data", in *SPIE Image and Video Processing II*, vol. 2182, (San Jose, USA), Feb. 1994.
- [3] M. M. Chang et al., "An algorithm for simultaneous motion estimation and scene segmentation", in *Proc. of the IEEE Int. Conf. on Acous., Speech, and Signal Proces.*, vol. V, (Adelaide, Australia), pp. 221-224, Apr. 1994.
- [4] M. T. Orchard, "Prediction motion field segmentation for image sequence coding", *IEEE Trans. Circuits and Systems for Video Technology*, vol. 3, no. 1, pp. 54-70, Feb. 1993.
- [5] C. Deutsch, A. Zaccarin and M. T. Orchard, "An estimation theoretic approach for robust predictive motion field segmentation", in *SPIE Image and Video Compression*, vol. 2186, (San Jose, USA), pp. 210-221, Jan. 1994.
- [6] C. Bouman and B. Liu, "Multiple resolution segmentation of textured images", *IEEE Trans. on Pattern Anal. and Machine Intel.*, pp. 99-113, vol. 13, no 2, Feb. 1991.
- [7] B. K. Horn and B. Schunck, "Determining optical flow", *Artificial Intelligence*, vol. 17, pp. 185-204, 1981.

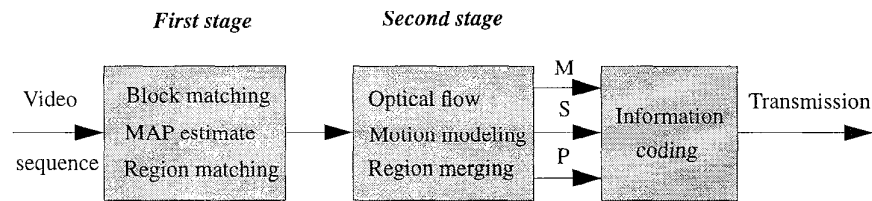


Figure 1: Block diagram of the proposed object-oriented algorithm

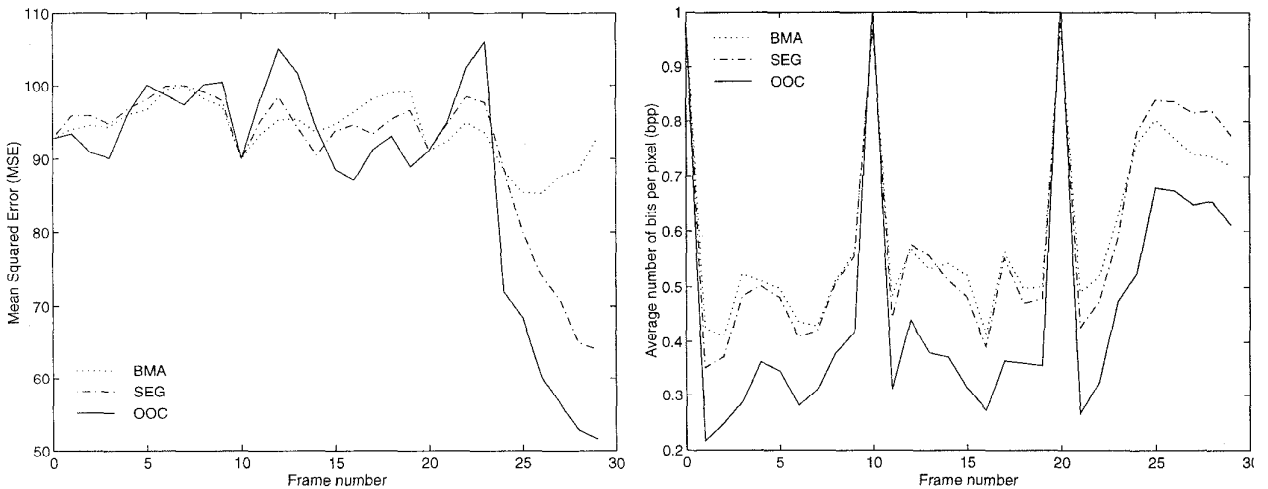


Figure 2: MSE and average number of bits per pixel for the *Table-tennis* sequence coded using the BMA, SEG and OOC algorithms

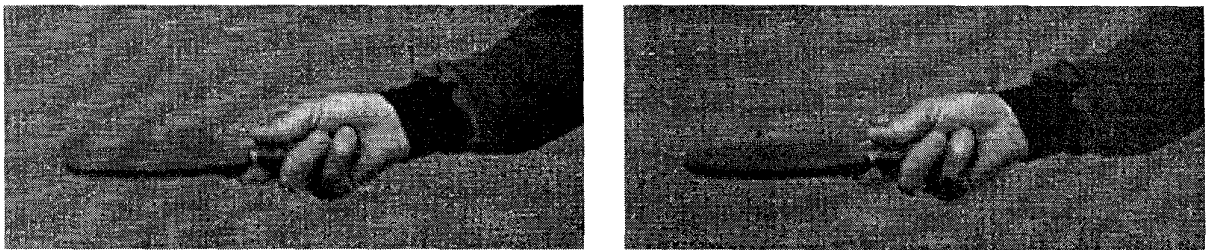


Figure 3: Motion compensated predicted image using block matching (left) and the proposed object-oriented coder (right)