Hierarchical Global Motion Estimation/Compensation in Low Bitrate Video Coding

Chung-Tao Chu, Dimitris Anastassiou, Shih-Fu Chang

Image and Advanced Television Laboratory
Department of Electrical Engineering, Columbia University
New York, N.Y. 10027 USA

Abstract

In this paper, we develop a hierarchical global motion estimation/compensation system which is used to improve the block-based motion compensation in the current low bitrate video coding standard H.263. The system first detects the background motion by examining the local motion information using hierarchical block matching and global motion estimation algorithm. If the background motion exists, the global motion compensation (**GMC**) is applied to the image before the H.263 block-based coding. Experiments are done under the typical low bitrate video conditions: low resolution and low frame rate. The results show that prediction errors on both motion vectors and texture are reduced. For sequences containing lots of background motion, the proposed system shows savings of up to 30% in total bitrate.

1. Introduction

High compression ratio video coding has been addressed in ITU-T SG 15/1, which targets video coding for low bitrate communication. It has outlined nearterm specifications for audio-visual coding. The nearterm video coding algorithm H.263 [1] has been optimized in motion prediction and frame formats. Motion estimation is done by block matching Motion vectors are DPCM coded, and motion-compensated predictive coding is used to generate the difference pictures, which are then transform coded using DCT.

One of the features of block-matching motion-compensated video coding at low bitrate is that overhead information usually takes a large portion of its bitstream. Especially, the motion information can take up to 25% of the bitrate, even though the motion vector is also DPCM coded. It is, then, important to reduce the motion information whenever possible.

When background motion exists, it is more likely that motion vector prediction generates errors resulting in more bits used for coding the motion vectors. This happens mostly on the border of the image where the macroblock motion is inconsistent with the other macroblocks inside the image (see figure 1). This issue

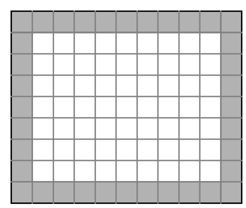


Figure 1: Macroblocks (gray area) **that tend to generate motion vector prediction error.**

becomes more important in low bitrate video coding because the image format is usually smaller, which means that the border macroblocks occupy a higher percentage of the image. Further, the low frame rate, which means large frame difference, results in relatively large background motion and increases the prediction error around the border. The frame distance becomes even

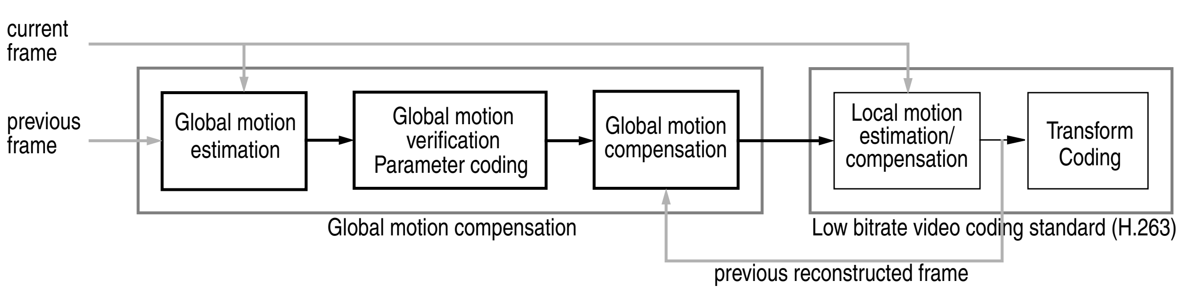
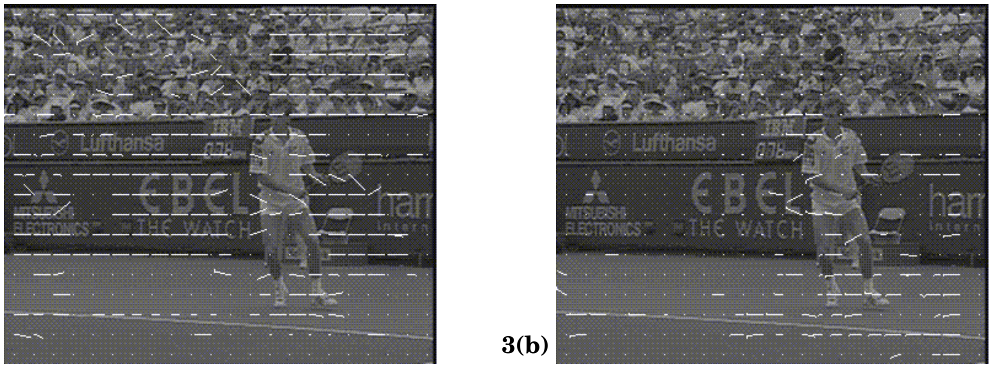


Figure 2: Block diagram of the global motion estimation/compensation.



3(a) 3(b)

Figure 3: Local motion vectors generated by H.263. (pictures from “Stefan” sequence)
3(a) without global motion compensation.
3(b) with global motion compensation.

larger between two P frames if any B-type picture is used.

Inconsistent motion prediction also happens in regions where block matching generates ambiguous results. This is because block matching is optimized to find the minimal block difference instead of true motion. Thus, the local minimum will take over the true motion and contribute to more prediction errors.

Taking into account the above considerations, global motion compensation (see figure 2) can help align the vector field and reduce the prediction error. Figure 3(a) shows the macroblock motion generated by H.263 and figure 3(b) shows the much smoother and more consistent macroblock motions after global motion compensation. Global motion estimation algorithms have been proposed in [2,3,4,5,6]. These algorithms differ in the model to represent the camera motion, as well as in the technique to estimate the parameters of the chosen model. However, in low bitrate video coding with characteristics of low resolution and low frame rate, these models become inefficient to provide good motion prediction, and those algorithms need to be modified for better motion estimation. Here, we choose a more general model that can describe more types of motion. Also we provide an algorithm to estimate the parameters for this motion model [10].

2. Global Motion Compensation

A video sequence can be more efficiently coded if its global motion is compensated first. Even a standard video codec like H.263 can benefit from such technique when there are large and dominant global motions. In [2,4,5] a model of four parameters corresponding to panning and zooming is used. In [3], 2-D rotation is considered and the model uses six parameters. In [6], an affine model with six parameters simplified from an eight parameters model is used. This affine model cannot fully describe 3-D rotation since it uses orthographic projection model and preserves parallelism, and is actually the same as that in [3].

These models are not sufficient when the global motion occurs on a scene with changing depth. Moreover, low bitrate video usually has low frame rate and large frame distance. Large frame distance implies that the global motion can be a combination of several types and the resulting motion may be more complicated. In order to obtain a better prediction on the background motion, a set of eight parameters [7,8] describing the mapping between two images is used to model the global motion (1):

$$S_{f_n}(X, Y) = S_{f_{n-1}}\left(\frac{a_1X + a_2Y + a_3}{a_7X + a_8Y + 1}, \frac{a_4X + a_5Y + a_6}{a_7X + a_8Y + 1}\right) \quad (1)$$

where $S_{fn}(x, y)$ is the pixel value at coordinate (x, y) in frame f_n. This model is capable of describing 3-D translation and 3-D rotation, thus all types of camera motions such as panning, zooming, rotation, and their combinations are included. If background motion exists, the image will be modified using equation (1). Then, the modified image is used as the reference of the H.263 block-based coding.

3. Global Motion Estimation

Given the motion model, the global motion estimation is a problem of least squares minimizations of the motion parameters. There are three techniques for solving this problem. The first one is a differential technique [2,3,7], which uses the information of local gradient and frame difference to find the motion parameters. However, in low bitrate video, it suffers from noises and large motions which happen more often because of larger frame distances. The second one is a matching technique [4,6], which is a generalized block matching that is carried on the background. For a motion model with more parameters, it may be very time consuming in

order to find the best match. The third method is a regression technique [5]. It first estimates the local motion and then uses the local information to find a global motion that minimizes the least square error. Refinements are carried out during the estimation to eliminate noise. In order to reduce the computational load as well as to increase the estimation robustness and accuracy, we propose a hierarchical motion estimation/compensation [10] similar to the third technique, based on the motion model (1). While full-search block matching is widely used to find the minimal block difference, hierarchical block matching, whose result usually approximates true motion, is used in the proposed algorithm. Further, hierarchical motion compensation is used to find the motion parameters with increasing accuracy at each step.

4. Hierarchical Global Motion Estimation/ Compensation

The algorithm is described in figure 4.

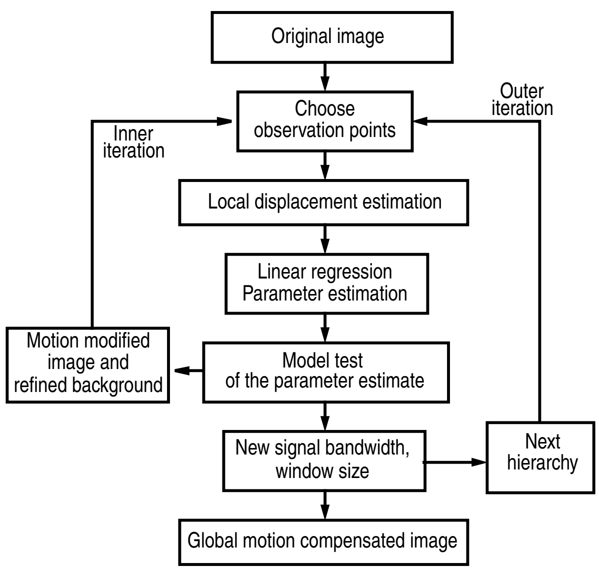


Figure 4. Global motion estimation.

First, equation (1) is re-written into a pair of linear equations. This is a linear system of dimension 8.

$$\begin{cases} X' = a_1X + a_2Y + a_3 - a_7XX' - a_8YX' \\ Y' = a_4X + a_5Y + a_6 - a_7XY' - a_8YY' \end{cases} \qquad (2)$$

Second, observation points are chosen throughout the image background and their local motions (displacements from (X, Y) to (X′,Y′)) are estimated. Ideal observation points should be chosen in image areas with higher activity. The local displacements are calculated using hierarchical block matching [10], because it is necessary to estimate the true motion, rather than only minimize displaced frame differences. The estimation uses different measurement window sizes, signal bandwidths, and maximum update displacements on different hierarchical levels. With the hierarchical structure, it is easy to modify the searching range/step for larger motion with few increase in computational load.

Then, the collection of observation points and their local motion information are substituted into the linear equations (2) to solve the motion parameter sets (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8). Four observation points are enough to solve the equations. More observation points and linear regression are introduced to minimize the estimation error. Refinement is carried out during the estimation to eliminate noisy observation points.

After each motion parameter estimation, a model test is performed to modify the image based on the estimated motion. The background region is also refined by exploiting the pixels that do not support the motion estimate. The next iteration of motion estimation is then carried out on the modified image and background region. These inner iterations will continue until no further improvement can be achieved, i.e. a convergent motion is reached.

Hierarchical motion compensation is combined with hierarchical block-matching. That is, for each hierarchy with local motion estimated using different window sizes and signal bandwidths, the global motion is estimated with different accuracy and the motion compensation is carried out. In our experiment, three hierarchies (outer iterations) are used.

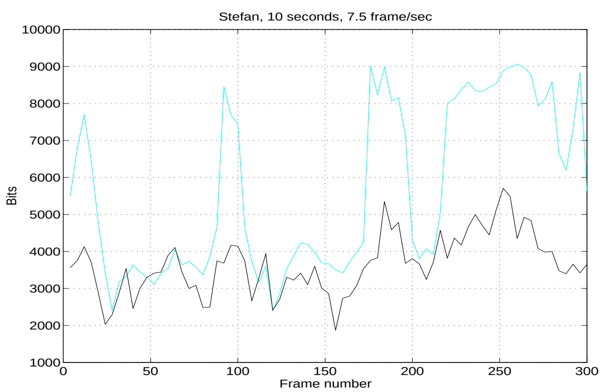
The presented motion estimation features low computing complexity compared with the matching technique. In our experiments, it requires the same order of computational load as those required in H.263 block-based motion estimation. Further, it differs from the mathematical differential model, which solves the linear system using indirect information (local gradient and difference). By using direct information (local displacement), it requires much fewer iterations to converge to a stable solution for the linear system (Eq. (2)).

5. Experimental Results

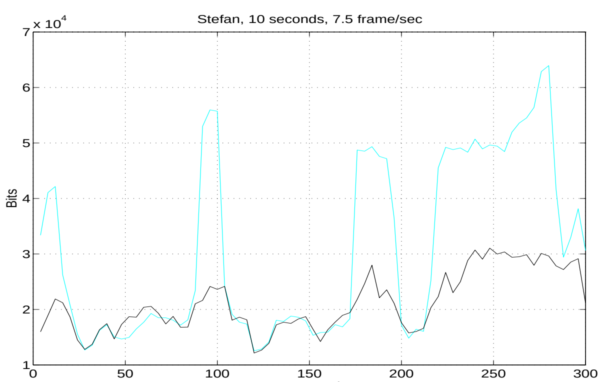
In the real applications, since global motion introduces extra overhead, a verification test is used to evaluate the global motion. Since global motion compensation is more helpful if the motion dominates the image, we use frame difference as the criterion in the verification process. Comparing with the current frame, if the global-motion compensated image has a frame difference of certain threshold less then that of the previ-

ous original frame, the global motion is adopted. Otherwise, only local block-motion compensation is used.

Simulations are performed on "*Stefan*" sequence in CIF format. The international standard H.263 codec is carried out at 7.5 frames/sec with fixed quantization. Experiment results are compared between **H.263** and **GMC+H.263**. Figure 5(a) compares the result of motion information and figure 5(b) compares the total bits coded for each frame. The high peaks of the H.263 curves represent the existence of large background motion. When global motion compensation is used, the motion information can be reduced by as much as 50%. The overall bitrate [**H.263** / **GMC+H.263**] is 233.4/ 163.4 kbps, which is a reduction of 30%.



5(a) comparison of coded motion information



5(b) comparison of total bitrate

Figure 5: Comparison of the total bitrate and vector information coded between H.263: ▬▬ and GMC+H.263: ▬▬.

6. Conclusion

A hierarchical global motion estimation/compensation is proposed for low bitrate video coding. The global motion parameters are estimated by using local motion information within the background region. Experiments are done with the standard codec H.263. The results have shown that GMC helps H.263 in reducing motion information as well as overall bitrate when there are large global motions.

References

1. ITU-T SGXV, Expert's Group for Very Low Bitrate Videophone, LBC-95, "Draft Recommendation H.263", July 1995.
2, M.Hotter, "Differential estimation of the global motion parameters zoom and pan", Signal Processing: Image Communication, vol. 16, pp. 249-265, March 1989
3. S.F. Wu and J. Kittler, "A differential method for simultaneous estimation of rotation, change of scale and translation", Signal Processing: Image communication, vol. 2, no. 1, pp. 69-80, May 1990.
4. D. Adolph and R. Buschmann, "1.15 Mbit/s coding of video signals including global motion compensation", Signal Processing: Image communication, vol. 3, nos. 2-3, pp. 259-274, June 1991.
5. Y.T. Tse and R.L. Baker, "Global zoom/pan estimation and compensation for video compression", IEEE Proc. ICASSP'91, vol. IV, pp. 2725-2728, Toronto, Canada, May 1991.
6. F. Moscheni, F. Dufaux and M. Kunt, "A new two-stage global/local motion estimation based on a background/foreground segmentation", IEEE Proc. ICASSP'95, vol. IV, pp. 2261-2265, 1995.
7. P. Anandan, J.R. Bergen, K.J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation", Motion analysis and Image Sequence Processing, pp. 1-22. Kluwer Academic Publishers, 1993.
8. T.S.Huang, A.N.Netravali, "Motion and Structure from Feature Correspondences: A Review", Proceedings of IEEE, Vol.82, No.2, February. 1994.
9. M. Bierling, "Displacement estimation by hierarchical block matching", SPIE, Visual Communication and Image Processing'88, vol.1001, pp. 942-951, 1988.
10. C.-T. Chu, "Motion Compensation at VOP Level", M0871, MPEG'96.

Hierarchical Global Motion Estimation/Compensation in Low Bitrate Video Coding

Chung-Tao Chu, Dimitris Anastassiou, Shih-Fu Chang

Image and Advanced Television Laboratory
Department of Electrical Engineering, Columbia University
New York, N.Y. 10027 USA

Abstract

In this paper, we develop a hierarchical global motion estimation/compensation system which is used to improve the block-based motion compensation in the current low bitrate video coding standard H.263. The system first detects the background motion by examining the local motion information using hierarchical block matching and global motion estimation algorithm. If the background motion exists, the global motion compensation (**GMC**) is applied to the image before the H.263 block-based coding. Experiments are done under the typical low bitrate video conditions: low resolution and low frame rate. The results show that prediction errors on both motion vectors and texture are reduced. For sequences containing lots of background motion, the proposed system shows savings of up to 30% in total bitrate.

1. Introduction

High compression ratio video coding has been addressed in ITU-T SG 15/1, which targets video coding for low bitrate communication. It has outlined nearterm specifications for audio-visual coding. The nearterm video coding algorithm H.263 [1] has been optimized in motion prediction and frame formats. Motion estimation is done by block matching Motion vectors are DPCM coded, and motion-compensated predictive coding is used to generate the difference pictures, which are then transform coded using DCT.

One of the features of block-matching motion-compensated video coding at low bitrate is that overhead information usually takes a large portion of its bitstream. Especially, the motion information can take up to 25% of the bitrate, even though the motion vector is also DPCM coded. It is, then, important to reduce the motion information whenever possible.

When background motion exists, it is more likely that motion vector prediction generates errors resulting in more bits used for coding the motion vectors. This happens mostly on the border of the image where the macroblock motion is inconsistent with the other macroblocks inside the image (see figure 1). This issue

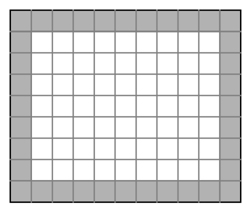


Figure 1: Macroblocks (gray area) **that tend to generate motion vector prediction error.**

becomes more important in low bitrate video coding because the image format is usually smaller, which means that the border macroblocks occupy a higher percentage of the image. Further, the low frame rate, which means large frame difference, results in relatively large background motion and increases the prediction error around the border. The frame distance becomes even

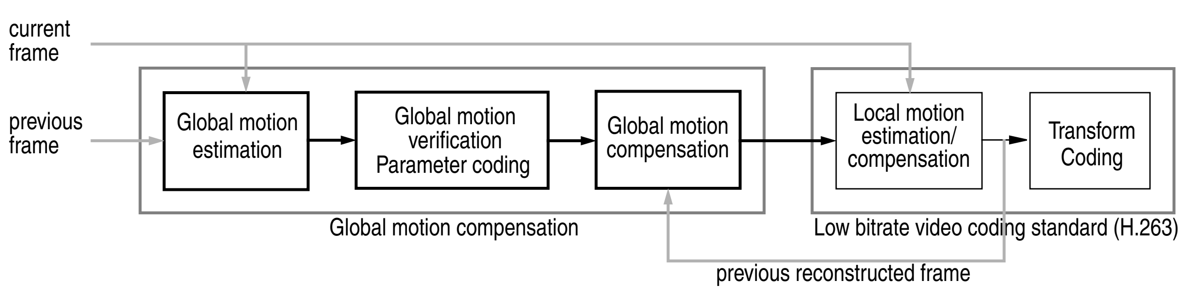


Figure 2: Block diagram of the global motion estimation/compensation.

3(a)  3(b) 

Figure 3: Local motion vectors generated by H.263. (pictures from "Stefan" sequence)
3(a) without global motion compensation.
3(b) with global motion compensation.

larger between two P frames if any B-type picture is used.

Inconsistent motion prediction also happens in regions where block matching generates ambiguous results. This is because block matching is optimized to find the minimal block difference instead of true motion. Thus, the local minimum will take over the true motion and contribute to more prediction errors.

Taking into account the above considerations, global motion compensation (see figure 2) can help align the vector field and reduce the prediction error. Figure 3(a) shows the macroblock motion generated by H.263 and figure 3(b) shows the much smoother and more consistent macroblock motions after global motion compensation. Global motion estimation algorithms have been proposed in [2,3,4,5,6]. These algorithms differ in the model to represent the camera motion, as well as in the technique to estimate the parameters of the chosen model. However, in low bitrate video coding with characteristics of low resolution and low frame rate, these models become inefficient to provide good motion prediction, and those algorithms need to be modified for better motion estimation. Here, we choose a more general model that can describe more types of motion. Also we provide an algorithm to estimate the parameters for this motion model [10].

2. Global Motion Compensation

A video sequence can be more efficiently coded if its global motion is compensated first. Even a standard video codec like H.263 can benefit from such technique when there are large and dominant global motions. In [2,4,5] a model of four parameters corresponding to panning and zooming is used. In [3], 2-D rotation is considered and the model uses six parameters. In [6], an affine model with six parameters simplified from an eight parameters model is used. This affine model cannot fully describe 3-D rotation since it uses orthographic projection model and preserves parallelism, and is actually the same as that in [3].

These models are not sufficient when the global motion occurs on a scene with changing depth. Moreover, low bitrate video usually has low frame rate and large frame distance. Large frame distance implies that the global motion can be a combination of several types and the resulting motion may be more complicated. In order to obtain a better prediction on the background motion, a set of eight parameters [7,8] describing the mapping between two images is used to model the global motion (1):

$$S_{f_n}(X, Y) = S_{f_{n-1}}\left(\frac{a_1X + a_2Y + a_3}{a_7X + a_8Y + 1}, \frac{a_4X + a_5Y + a_6}{a_7X + a_8Y + 1}\right) \quad (1)$$

where S_{fn}(x, y) is the pixel value at coordinate (x, y) in frame f_n. This model is capable of describing 3-D translation and 3-D rotation, thus all types of camera motions such as panning, zooming, rotation, and their combinations are included. If background motion exists, the image will be modified using equation (1). Then, the modified image is used as the reference of the H.263 block-based coding.

3. Global Motion Estimation

Given the motion model, the global motion estimation is a problem of least squares minimizations of the motion parameters. There are three techniques for solving this problem. The first one is a differential technique [2,3,7], which uses the information of local gradient and frame difference to find the motion parameters. However, in low bitrate video, it suffers from noises and large motions which happen more often because of larger frame distances. The second one is a matching technique [4,6], which is a generalized block matching that is carried on the background. For a motion model with more parameters, it may be very time consuming in

order to find the best match. The third method is a regression technique [5]. It first estimates the local motion and then uses the local information to find a global motion that minimizes the least square error. Refinements are carried out during the estimation to eliminate noise. In order to reduce the computational load as well as to increase the estimation robustness and accuracy, we propose a hierarchical motion estimation/compensation [10] similar to the third technique, based on the motion model (1). While full-search block matching is widely used to find the minimal block difference, hierarchical block matching, whose result usually approximates true motion, is used in the proposed algorithm. Further, hierarchical motion compensation is used to find the motion parameters with increasing accuracy at each step.

4. Hierarchical Global Motion Estimation/ Compensation

The algorithm is described in figure 4.

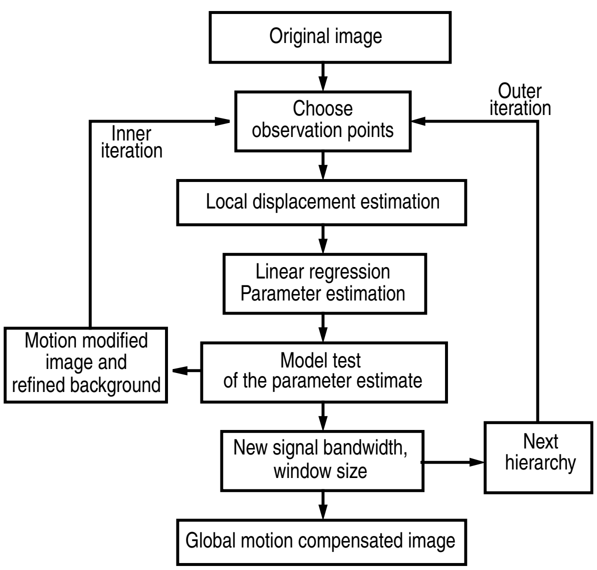


Figure 4. Global motion estimation.

First, equation (1) is re-written into a pair of linear equations. This is a linear system of dimension 8.

$$\begin{cases} X' = a_1X + a_2Y + a_3 - a_7XX' - a_8YX' \\ Y' = a_4X + a_5Y + a_6 - a_7XY' - a_8YY' \end{cases} \quad (2)$$

Second, observation points are chosen throughout the image background and their local motions (displacements from (X, Y) to (X′,Y′)) are estimated. Ideal observation points should be chosen in image areas with higher activity. The local displacements are calculated using hierarchical block matching [10], because it is necessary to estimate the true motion, rather than only minimize displaced frame differences. The estimation uses different measurement window sizes, signal bandwidths, and maximum update displacements on different hierarchical levels. With the hierarchical structure, it is easy to modify the searching range/step for larger motion with few increase in computational load.

Then, the collection of observation points and their local motion information are substituted into the linear equations (2) to solve the motion parameter sets (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8). Four observation points are enough to solve the equations. More observation points and linear regression are introduced to minimize the estimation error. Refinement is carried out during the estimation to eliminate noisy observation points.

After each motion parameter estimation, a model test is performed to modify the image based on the estimated motion. The background region is also refined by exploiting the pixels that do not support the motion estimate. The next iteration of motion estimation is then carried out on the modified image and background region. These inner iterations will continue until no further improvement can be achieved, i.e. a convergent motion is reached.

Hierarchical motion compensation is combined with hierarchical block-matching. That is, for each hierarchy with local motion estimated using different window sizes and signal bandwidths, the global motion is estimated with different accuracy and the motion compensation is carried out. In our experiment, three hierarchies (outer iterations) are used.

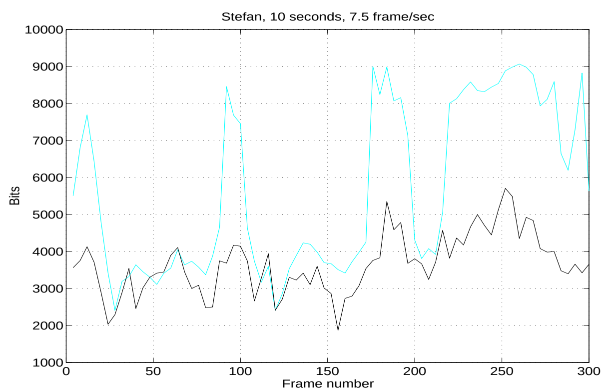
The presented motion estimation features low computing complexity compared with the matching technique. In our experiments, it requires the same order of computational load as those required in H.263 block-based motion estimation. Further, it differs from the mathematical differential model, which solves the linear system using indirect information (local gradient and difference). By using direct information (local displacement), it requires much fewer iterations to converge to a stable solution for the linear system (Eq. (2)).

5. Experimental Results

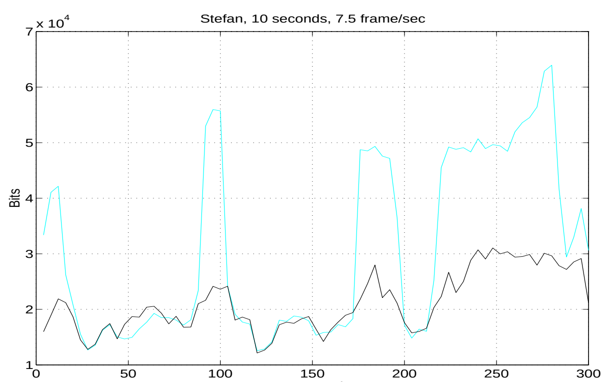
In the real applications, since global motion introduces extra overhead, a verification test is used to evaluate the global motion. Since global motion compensation is more helpful if the motion dominates the image, we use frame difference as the criterion in the verification process. Comparing with the current frame, if the global-motion compensated image has a frame difference of certain threshold less then that of the previ-

ous original frame, the global motion is adopted. Otherwise, only local block-motion compensation is used.

Simulations are performed on "*Stefan*" sequence in CIF format. The international standard H.263 codec is carried out at 7.5 frames/sec with fixed quantization. Experiment results are compared between **H.263** and **GMC+H.263**. Figure 5(a) compares the result of motion information and figure 5(b) compares the total bits coded for each frame. The high peaks of the H.263 curves represent the existence of large background motion. When global motion compensation is used, the motion information can be reduced by as much as 50%. The overall bitrate [**H.263** / **GMC+H.263**] is 233.4/163.4 kbps, which is a reduction of 30%.



5(a) comparison of coded motion information



5(b) comparison of total bitrate

Figure 5: Comparison of the total bitrate and vector information coded between H.263: ▬ and GMC+H.263: ▬.

6. Conclusion

A hierarchical global motion estimation/compensation is proposed for low bitrate video coding. The global motion parameters are estimated by using local motion information within the background region. Experiments are done with the standard codec H.263. The results have shown that GMC helps H.263 in reducing motion information as well as overall bitrate when there are large global motions.

References

1. ITU-T SGXV, Expert's Group for Very Low Bitrate Videophone, LBC-95, "Draft Recommendation H.263", July 1995.
2, M.Hotter, "Differential estimation of the global motion parameters zoom and pan", Signal Processing: Image Communication, vol. 16, pp. 249-265, March 1989
3. S.F. Wu and J. Kittler, "A differential method for simultaneous estimation of rotation, change of scale and translation", Signal Processing: Image communication, vol. 2, no. 1, pp. 69-80, May 1990.
4. D. Adolph and R. Buschmann, "1.15 Mbit/s coding of video signals including global motion compensation", Signal Processing: Image communication, vol. 3, nos. 2-3, pp. 259-274, June 1991.
5. Y.T. Tse and R.L. Baker, "Global zoom/pan estimation and compensation for video compression", IEEE Proc. ICASSP'91, vol. IV, pp. 2725-2728, Toronto, Canada, May 1991.
6. F. Moscheni, F. Dufaux and M. Kunt, "A new two-stage global/local motion estimation based on a background/foreground segmentation", IEEE Proc. ICASSP'95, vol. IV, pp. 2261-2265, 1995.
7. P. Anandan, J.R. Bergen, K.J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation", Motion analysis and Image Sequence Processing, pp. 1-22. Kluwer Academic Publishers, 1993.
8. T.S.Huang, A.N.Netravali, "Motion and Structure from Feature Correspondences: A Review", Proceedings of IEEE, Vol.82, No.2, February. 1994.
9. M. Bierling, "Displacement estimation by hierarchical block matching", SPIE, Visual Communication and Image Processing'88, vol.1001, pp. 942-951, 1988.
10. C.-T. Chu, "Motion Compensation at VOP Level", M0871, MPEG'96.

Hierarchical Global Motion Estimation/Compensation in Low Bitrate Video Coding

Chung-Tao Chu, Dimitris Anastassiou, Shih-Fu Chang

Image and Advanced Television Laboratory
Department of Electrical Engineering, Columbia University
New York, N.Y. 10027 USA

Abstract

In this paper, we develop a hierarchical global motion estimation/compensation system which is used to improve the block-based motion compensation in the current low bitrate video coding standard H.263. The system first detects the background motion by examining the local motion information using hierarchical block matching and global motion estimation algorithm. If the background motion exists, the global motion compensation (**GMC**) is applied to the image before the H.263 block-based coding. Experiments are done under the typical low bitrate video conditions: low resolution and low frame rate. The results show that prediction errors on both motion vectors and texture are reduced. For sequences containing lots of background motion, the proposed system shows savings of up to 30% in total bitrate.

1. Introduction

High compression ratio video coding has been addressed in ITU-T SG 15/1, which targets video coding for low bitrate communication. It has outlined nearterm specifications for audio-visual coding. The nearterm video coding algorithm H.263 [1] has been optimized in motion prediction and frame formats. Motion estimation is done by block matching Motion vectors are DPCM coded, and motion-compensated predictive coding is used to generate the difference pictures, which are then transform coded using DCT.

One of the features of block-matching motion-compensated video coding at low bitrate is that overhead information usually takes a large portion of its bitstream. Especially, the motion information can take up to 25% of the bitrate, even though the motion vector is also DPCM coded. It is, then, important to reduce the motion information whenever possible.

When background motion exists, it is more likely that motion vector prediction generates errors resulting in more bits used for coding the motion vectors. This happens mostly on the border of the image where the macroblock motion is inconsistent with the other macroblocks inside the image (see figure 1). This issue

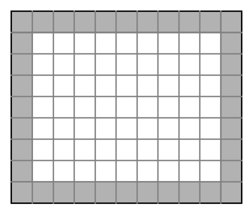


Figure 1: Macroblocks (gray area) **that tend to generate motion vector prediction error.**

becomes more important in low bitrate video coding because the image format is usually smaller, which means that the border macroblocks occupy a higher percentage of the image. Further, the low frame rate, which means large frame difference, results in relatively large background motion and increases the prediction error around the border. The frame distance becomes even

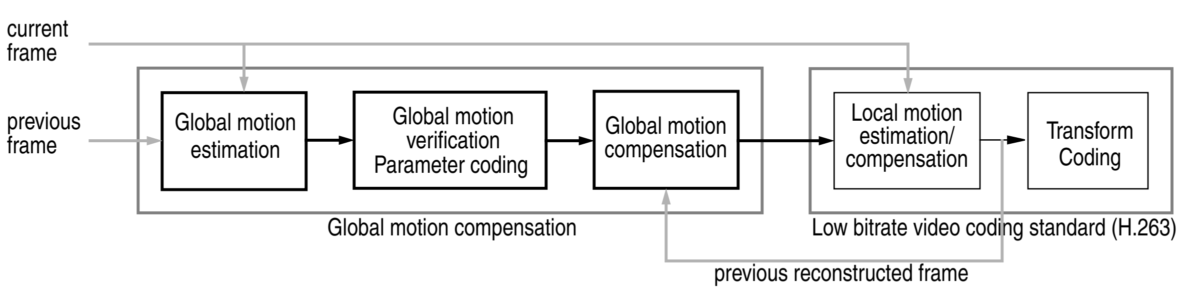
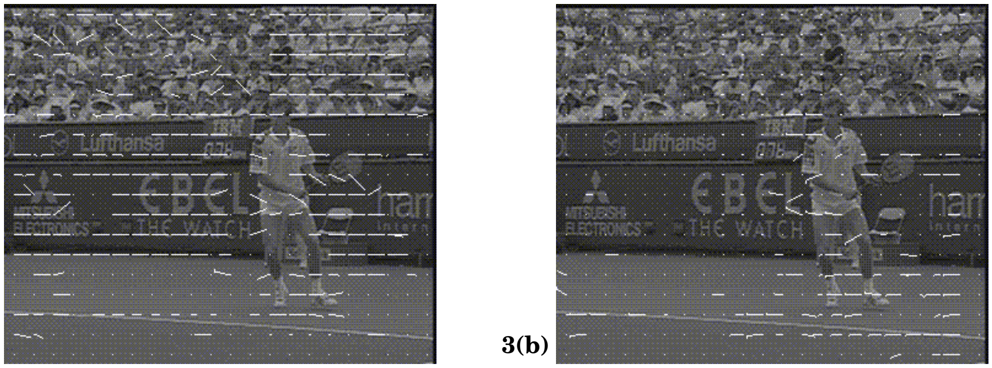


Figure 2: Block diagram of the global motion estimation/compensation.



3(a) 3(b)

Figure 3: Local motion vectors generated by H.263. (pictures from "Stefan" sequence)
3(a) without global motion compensation.
3(b) with global motion compensation.

larger between two P frames if any B-type picture is used.

Inconsistent motion prediction also happens in regions where block matching generates ambiguous results. This is because block matching is optimized to find the minimal block difference instead of true motion. Thus, the local minimum will take over the true motion and contribute to more prediction errors.

Taking into account the above considerations, global motion compensation (see figure 2) can help align the vector field and reduce the prediction error. Figure 3(a) shows the macroblock motion generated by H.263 and figure 3(b) shows the much smoother and more consistent macroblock motions after global motion compensation. Global motion estimation algorithms have been proposed in [2,3,4,5,6]. These algorithms differ in the model to represent the camera motion, as well as in the technique to estimate the parameters of the chosen model. However, in low bitrate video coding with characteristics of low resolution and low frame rate, these models become inefficient to provide good motion prediction, and those algorithms need to be modified for better motion estimation. Here, we choose a more general model that can describe more types of motion. Also we provide an algorithm to estimate the parameters for this motion model [10].

2. Global Motion Compensation

A video sequence can be more efficiently coded if its global motion is compensated first. Even a standard video codec like H.263 can benefit from such technique when there are large and dominant global motions. In [2,4,5] a model of four parameters corresponding to panning and zooming is used. In [3], 2-D rotation is considered and the model uses six parameters. In [6], an affine model with six parameters simplified from an eight parameters model is used. This affine model cannot fully describe 3-D rotation since it uses orthographic projection model and preserves parallelism, and is actually the same as that in [3].

These models are not sufficient when the global motion occurs on a scene with changing depth. Moreover, low bitrate video usually has low frame rate and large frame distance. Large frame distance implies that the global motion can be a combination of several types and the resulting motion may be more complicated. In order to obtain a better prediction on the background motion, a set of eight parameters [7,8] describing the mapping between two images is used to model the global motion (1):

$$S_{f_n}(X, Y) = S_{f_{n-1}}\left(\frac{a_1 X + a_2 Y + a_3}{a_7 X + a_8 Y + 1}, \frac{a_4 X + a_5 Y + a_6}{a_7 X + a_8 Y + 1}\right) \quad (1)$$

where $S_{fn}(x, y)$ is the pixel value at coordinate (x, y) in frame f_n. This model is capable of describing 3-D translation and 3-D rotation, thus all types of camera motions such as panning, zooming, rotation, and their combinations are included. If background motion exists, the image will be modified using equation (1). Then, the modified image is used as the reference of the H.263 block-based coding.

3. Global Motion Estimation

Given the motion model, the global motion estimation is a problem of least squares minimizations of the motion parameters. There are three techniques for solving this problem. The first one is a differential technique [2,3,7], which uses the information of local gradient and frame difference to find the motion parameters. However, in low bitrate video, it suffers from noises and large motions which happen more often because of larger frame distances. The second one is a matching technique [4,6], which is a generalized block matching that is carried on the background. For a motion model with more parameters, it may be very time consuming in

order to find the best match. The third method is a regression technique [5]. It first estimates the local motion and then uses the local information to find a global motion that minimizes the least square error. Refinements are carried out during the estimation to eliminate noise. In order to reduce the computational load as well as to increase the estimation robustness and accuracy, we propose a hierarchical motion estimation/compensation [10] similar to the third technique, based on the motion model (1). While full-search block matching is widely used to find the minimal block difference, hierarchical block matching, whose result usually approximates true motion, is used in the proposed algorithm. Further, hierarchical motion compensation is used to find the motion parameters with increasing accuracy at each step.

4. Hierarchical Global Motion Estimation/ Compensation

The algorithm is described in figure 4.

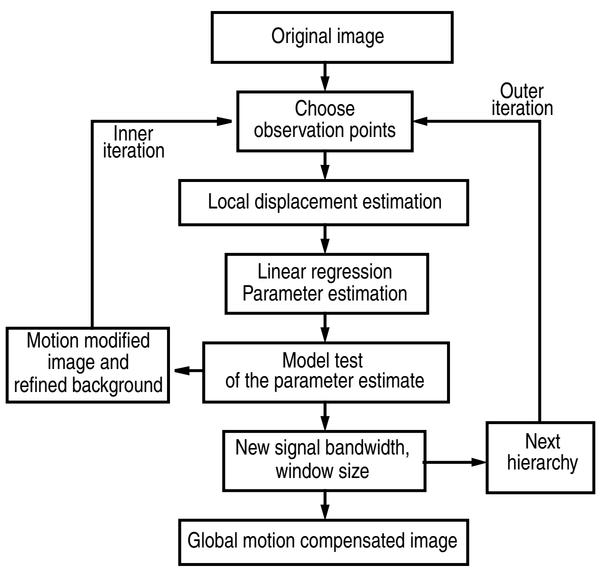


Figure 4. Global motion estimation.

First, equation (1) is re-written into a pair of linear equations. This is a linear system of dimension 8.

$$\begin{cases} X' = a_1 X + a_2 Y + a_3 - a_7 XX' - a_8 YX' \\ Y' = a_4 X + a_5 Y + a_6 - a_7 XY' - a_8 YY' \end{cases} \quad (2)$$

Second, observation points are chosen throughout the image background and their local motions (displacements from (X, Y) to (X′,Y′)) are estimated. Ideal observation points should be chosen in image areas with higher activity. The local displacements are calculated using hierarchical block matching [10], because it is necessary to estimate the true motion, rather than only minimize displaced frame differences. The estimation uses different measurement window sizes, signal bandwidths, and maximum update displacements on different hierarchical levels. With the hierarchical structure, it is easy to modify the searching range/step for larger motion with few increase in computational load.

Then, the collection of observation points and their local motion information are substituted into the linear equations (2) to solve the motion parameter sets (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8). Four observation points are enough to solve the equations. More observation points and linear regression are introduced to minimize the estimation error. Refinement is carried out during the estimation to eliminate noisy observation points.

After each motion parameter estimation, a model test is performed to modify the image based on the estimated motion. The background region is also refined by exploiting the pixels that do not support the motion estimate. The next iteration of motion estimation is then carried out on the modified image and background region. These inner iterations will continue until no further improvement can be achieved, i.e. a convergent motion is reached.

Hierarchical motion compensation is combined with hierarchical block-matching. That is, for each hierarchy with local motion estimated using different window sizes and signal bandwidths, the global motion is estimated with different accuracy and the motion compensation is carried out. In our experiment, three hierarchies (outer iterations) are used.

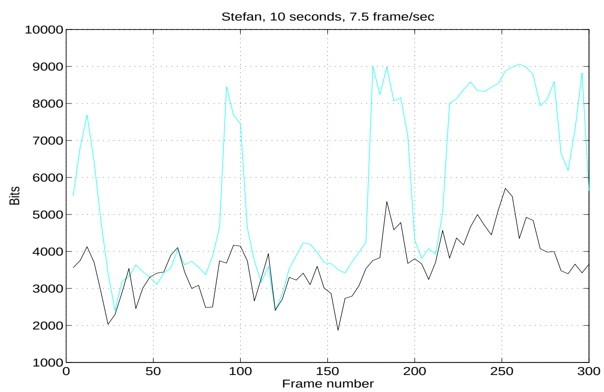
The presented motion estimation features low computing complexity compared with the matching technique. In our experiments, it requires the same order of computational load as those required in H.263 block-based motion estimation. Further, it differs from the mathematical differential model, which solves the linear system using indirect information (local gradient and difference). By using direct information (local displacement), it requires much fewer iterations to converge to a stable solution for the linear system (Eq. (2)).

5. Experimental Results

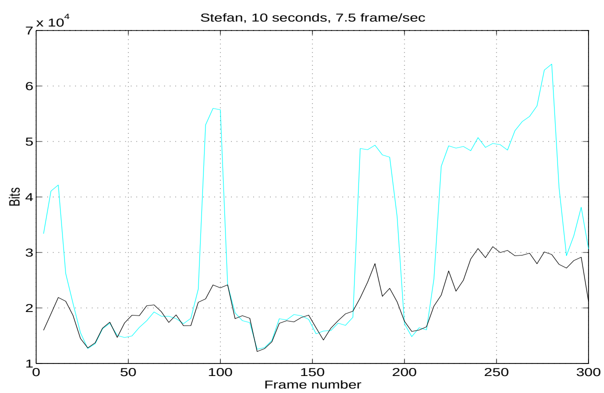
In the real applications, since global motion introduces extra overhead, a verification test is used to evaluate the global motion. Since global motion compensation is more helpful if the motion dominates the image, we use frame difference as the criterion in the verification process. Comparing with the current frame, if the global-motion compensated image has a frame difference of certain threshold less then that of the previ-

ous original frame, the global motion is adopted. Otherwise, only local block-motion compensation is used.

Simulations are performed on "*Stefan*" sequence in CIF format. The international standard H.263 codec is carried out at 7.5 frames/sec with fixed quantization. Experiment results are compared between **H.263** and **GMC+H.263**. Figure 5(a) compares the result of motion information and figure 5(b) compares the total bits coded for each frame. The high peaks of the H.263 curves represent the existence of large background motion. When global motion compensation is used, the motion information can be reduced by as much as 50%. The overall bitrate [**H.263** / **GMC+H.263**] is 233.4/ 163.4 kbps, which is a reduction of 30%.



5(a) comparison of coded motion information



5(b) comparison of total bitrate

Figure 5: Comparison of the total bitrate and vector information coded between H.263: —— and GMC+H.263: ——.

6. Conclusion

A hierarchical global motion estimation/compensation is proposed for low bitrate video coding. The global motion parameters are estimated by using local motion information within the background region. Experiments are done with the standard codec H.263. The results have shown that GMC helps H.263 in reducing motion information as well as overall bitrate when there are large global motions.

References

1. ITU-T SGXV, Expert's Group for Very Low Bitrate Videophone, LBC-95, "Draft Recommendation H.263", July 1995.
2, M.Hotter, "Differential estimation of the global motion parameters zoom and pan", Signal Processing: Image Communication, vol. 16, pp. 249-265, March 1989
3. S.F. Wu and J. Kittler, "A differential method for simultaneous estimation of rotation, change of scale and translation", Signal Processing: Image communication, vol. 2, no. 1, pp. 69-80, May 1990.
4. D. Adolph and R. Buschmann, "1.15 Mbit/s coding of video signals including global motion compensation", Signal Processing: Image communication, vol. 3, nos. 2-3, pp. 259-274, June 1991.
5. Y.T. Tse and R.L. Baker, "Global zoom/pan estimation and compensation for video compression", IEEE Proc. ICASSP'91, vol. IV, pp. 2725-2728, Toronto, Canada, May 1991.
6. F. Moscheni, F. Dufaux and M. Kunt, "A new two-stage global/local motion estimation based on a background/foreground segmentation", IEEE Proc. ICASSP'95, vol. IV, pp. 2261-2265, 1995.
7. P. Anandan, J.R. Bergen, K.J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation", Motion analysis and Image Sequence Processing, pp. 1-22. Kluwer Academic Publishers, 1993.
8. T.S.Huang, A.N.Netravali, "Motion and Structure from Feature Correspondences: A Review", Proceedings of IEEE, Vol.82, No.2, February. 1994.
9. M. Bierling, "Displacement estimation by hierarchical block matching", SPIE, Visual Communication and Image Processing'88, vol.1001, pp. 942-951, 1988.
10. C.-T. Chu, "Motion Compensation at VOP Level", M0871, MPEG'96.

Hierarchical Global Motion Estimation/Compensation in Low Bitrate Video Coding

Chung-Tao Chu, Dimitris Anastassiou, Shih-Fu Chang

Image and Advanced Television Laboratory
Department of Electrical Engineering, Columbia University
New York, N.Y. 10027 USA

Abstract

In this paper, we develop a hierarchical global motion estimation/compensation system which is used to improve the block-based motion compensation in the current low bitrate video coding standard H.263. The system first detects the background motion by examining the local motion information using hierarchical block matching and global motion estimation algorithm. If the background motion exists, the global motion compensation (**GMC**) is applied to the image before the H.263 block-based coding. Experiments are done under the typical low bitrate video conditions: low resolution and low frame rate. The results show that prediction errors on both motion vectors and texture are reduced. For sequences containing lots of background motion, the proposed system shows savings of up to 30% in total bitrate.

1. Introduction

High compression ratio video coding has been addressed in ITU-T SG 15/1, which targets video coding for low bitrate communication. It has outlined nearterm specifications for audio-visual coding. The nearterm video coding algorithm H.263 [1] has been optimized in motion prediction and frame formats. Motion estimation is done by block matching Motion vectors are DPCM coded, and motion-compensated predictive coding is used to generate the difference pictures, which are then transform coded using DCT.

One of the features of block-matching motion-compensated video coding at low bitrate is that overhead information usually takes a large portion of its bitstream. Especially, the motion information can take up to 25% of the bitrate, even though the motion vector is also DPCM coded. It is, then, important to reduce the motion information whenever possible.

When background motion exists, it is more likely that motion vector prediction generates errors resulting in more bits used for coding the motion vectors. This happens mostly on the border of the image where the macroblock motion is inconsistent with the other macroblocks inside the image (see figure 1). This issue

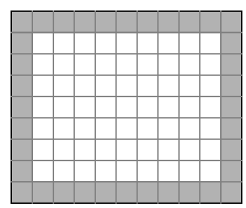


Figure 1: Macroblocks (gray area) **that tend to generate motion vector prediction error.**

becomes more important in low bitrate video coding because the image format is usually smaller, which means that the border macroblocks occupy a higher percentage of the image. Further, the low frame rate, which means large frame difference, results in relatively large background motion and increases the prediction error around the border. The frame distance becomes even

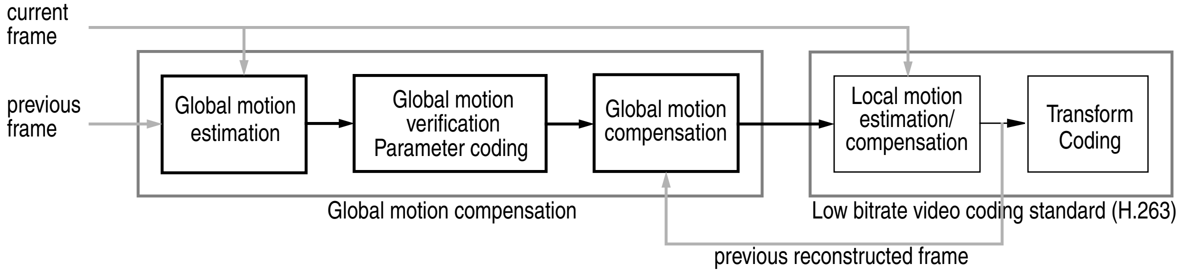


Figure 2: Block diagram of the global motion estimation/compensation.

3(a)  3(b) 

Figure 3: Local motion vectors generated by H.263. (pictures from "Stefan" sequence)
3(a) without global motion compensation.
3(b) with global motion compensation.

larger between two P frames if any B-type picture is used.

Inconsistent motion prediction also happens in regions where block matching generates ambiguous results. This is because block matching is optimized to find the minimal block difference instead of true motion. Thus, the local minimum will take over the true motion and contribute to more prediction errors.

Taking into account the above considerations, global motion compensation (see figure 2) can help align the vector field and reduce the prediction error. Figure 3(a) shows the macroblock motion generated by H.263 and figure 3(b) shows the much smoother and more consistent macroblock motions after global motion compensation. Global motion estimation algorithms have been proposed in [2,3,4,5,6]. These algorithms differ in the model to represent the camera motion, as well as in the technique to estimate the parameters of the chosen model. However, in low bitrate video coding with characteristics of low resolution and low frame rate, these models become inefficient to provide good motion prediction, and those algorithms need to be modified for better motion estimation. Here, we choose a more general model that can describe more types of motion. Also we provide an algorithm to estimate the parameters for this motion model [10].

2. Global Motion Compensation

A video sequence can be more efficiently coded if its global motion is compensated first. Even a standard video codec like H.263 can benefit from such technique when there are large and dominant global motions. In [2,4,5] a model of four parameters corresponding to panning and zooming is used. In [3], 2-D rotation is considered and the model uses six parameters. In [6], an affine model with six parameters simplified from an eight parameters model is used. This affine model cannot fully describe 3-D rotation since it uses orthographic projection model and preserves parallelism, and is actually the same as that in [3].

These models are not sufficient when the global motion occurs on a scene with changing depth. Moreover, low bitrate video usually has low frame rate and large frame distance. Large frame distance implies that the global motion can be a combination of several types and the resulting motion may be more complicated. In order to obtain a better prediction on the background motion, a set of eight parameters [7,8] describing the mapping between two images is used to model the global motion (1):

$$S_{f_n}(X, Y) = S_{f_{n-1}}\left(\frac{a_1X + a_2Y + a_3}{a_7X + a_8Y + 1}, \frac{a_4X + a_5Y + a_6}{a_7X + a_8Y + 1}\right) \quad (1)$$

where $S_{fn}(x, y)$ is the pixel value at coordinate (x, y) in frame f_n. This model is capable of describing 3-D translation and 3-D rotation, thus all types of camera motions such as panning, zooming, rotation, and their combinations are included. If background motion exists, the image will be modified using equation (1). Then, the modified image is used as the reference of the H.263 block-based coding.

3. Global Motion Estimation

Given the motion model, the global motion estimation is a problem of least squares minimizations of the motion parameters. There are three techniques for solving this problem. The first one is a differential technique [2,3,7], which uses the information of local gradient and frame difference to find the motion parameters. However, in low bitrate video, it suffers from noises and large motions which happen more often because of larger frame distances. The second one is a matching technique [4,6], which is a generalized block matching that is carried on the background. For a motion model with more parameters, it may be very time consuming in

order to find the best match. The third method is a regression technique [5]. It first estimates the local motion and then uses the local information to find a global motion that minimizes the least square error. Refinements are carried out during the estimation to eliminate noise. In order to reduce the computational load as well as to increase the estimation robustness and accuracy, we propose a hierarchical motion estimation/compensation [10] similar to the third technique, based on the motion model (1). While full-search block matching is widely used to find the minimal block difference, hierarchical block matching, whose result usually approximates true motion, is used in the proposed algorithm. Further, hierarchical motion compensation is used to find the motion parameters with increasing accuracy at each step.

4. Hierarchical Global Motion Estimation/ Compensation

The algorithm is described in figure 4.

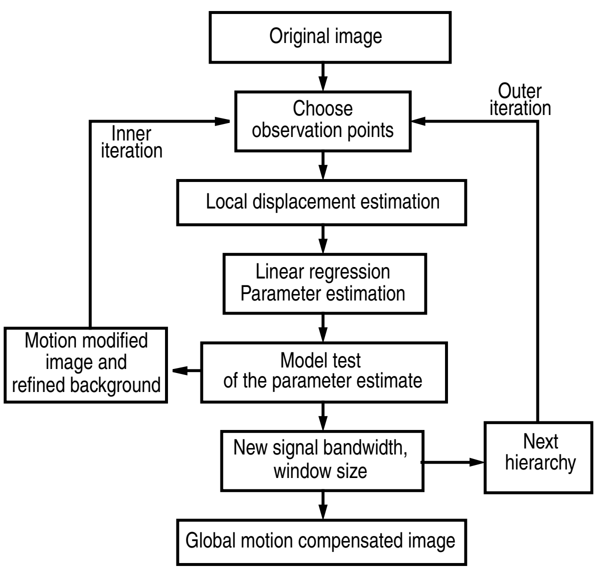


Figure 4. Global motion estimation.

First, equation (1) is re-written into a pair of linear equations. This is a linear system of dimension 8.

$$\begin{cases} X' = a_1X + a_2Y + a_3 - a_7XX' - a_8YX' \\ Y' = a_4X + a_5Y + a_6 - a_7XY' - a_8YY' \end{cases} \quad (2)$$

Second, observation points are chosen throughout the image background and their local motions (displacements from (X, Y) to (X′,Y′)) are estimated. Ideal observation points should be chosen in image areas with higher activity. The local displacements are calculated using hierarchical block matching [10], because it is necessary to estimate the true motion, rather than only minimize displaced frame differences. The estimation uses different measurement window sizes, signal bandwidths, and maximum update displacements on different hierarchical levels. With the hierarchical structure, it is easy to modify the searching range/step for larger motion with few increase in computational load.

Then, the collection of observation points and their local motion information are substituted into the linear equations (2) to solve the motion parameter sets (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8). Four observation points are enough to solve the equations. More observation points and linear regression are introduced to minimize the estimation error. Refinement is carried out during the estimation to eliminate noisy observation points.

After each motion parameter estimation, a model test is performed to modify the image based on the estimated motion. The background region is also refined by exploiting the pixels that do not support the motion estimate. The next iteration of motion estimation is then carried out on the modified image and background region. These inner iterations will continue until no further improvement can be achieved, i.e. a convergent motion is reached.

Hierarchical motion compensation is combined with hierarchical block-matching. That is, for each hierarchy with local motion estimated using different window sizes and signal bandwidths, the global motion is estimated with different accuracy and the motion compensation is carried out. In our experiment, three hierarchies (outer iterations) are used.

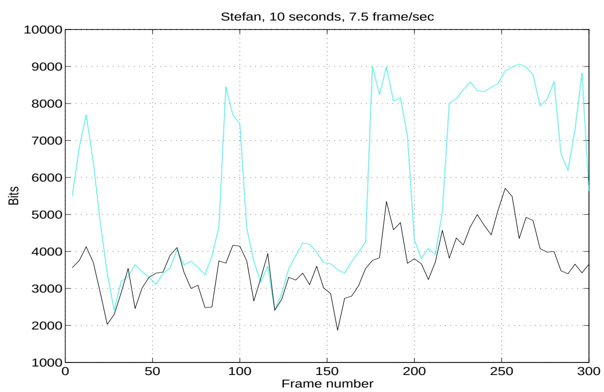
The presented motion estimation features low computing complexity compared with the matching technique. In our experiments, it requires the same order of computational load as those required in H.263 block-based motion estimation. Further, it differs from the mathematical differential model, which solves the linear system using indirect information (local gradient and difference). By using direct information (local displacement), it requires much fewer iterations to converge to a stable solution for the linear system (Eq. (2)).

5. Experimental Results

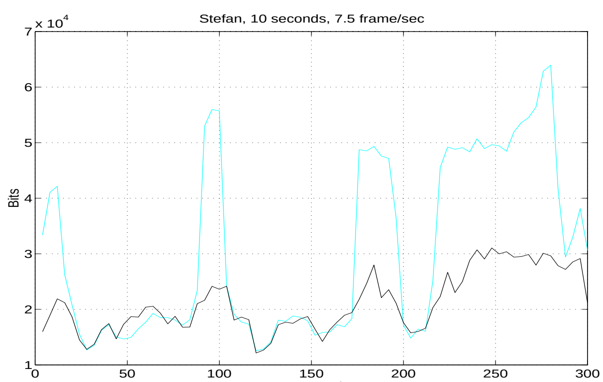
In the real applications, since global motion introduces extra overhead, a verification test is used to evaluate the global motion. Since global motion compensation is more helpful if the motion dominates the image, we use frame difference as the criterion in the verification process. Comparing with the current frame, if the global-motion compensated image has a frame difference of certain threshold less then that of the previ-

ous original frame, the global motion is adopted. Otherwise, only local block-motion compensation is used.

Simulations are performed on "*Stefan*" sequence in CIF format. The international standard H.263 codec is carried out at 7.5 frames/sec with fixed quantization. Experiment results are compared between **H.263** and **GMC+H.263**. Figure 5(a) compares the result of motion information and figure 5(b) compares the total bits coded for each frame. The high peaks of the H.263 curves represent the existence of large background motion. When global motion compensation is used, the motion information can be reduced by as much as 50%. The overall bitrate [**H.263** / **GMC+H.263**] is 233.4/ 163.4 kbps, which is a reduction of 30%.



5(a) comparison of coded motion information



5(b) comparison of total bitrate

Figure 5: Comparison of the total bitrate and vector information coded between H.263: ▬ and GMC+H.263: ▬.

6. Conclusion

A hierarchical global motion estimation/compensation is proposed for low bitrate video coding. The global motion parameters are estimated by using local motion information within the background region. Experiments are done with the standard codec H.263. The results have shown that GMC helps H.263 in reducing motion information as well as overall bitrate when there are large global motions.

References

1. ITU-T SGXV, Expert's Group for Very Low Bitrate Videophone, LBC-95, "Draft Recommendation H.263", July 1995.
2, M.Hotter, "Differential estimation of the global motion parameters zoom and pan", Signal Processing: Image Communication, vol. 16, pp. 249-265, March 1989
3. S.F. Wu and J. Kittler, "A differential method for simultaneous estimation of rotation, change of scale and translation", Signal Processing: Image communication, vol. 2, no. 1, pp. 69-80, May 1990.
4. D. Adolph and R. Buschmann, "1.15 Mbit/s coding of video signals including global motion compensation", Signal Processing: Image communication, vol. 3, nos. 2-3, pp. 259-274, June 1991.
5. Y.T. Tse and R.L. Baker, "Global zoom/pan estimation and compensation for video compression", IEEE Proc. ICASSP'91, vol. IV, pp. 2725-2728, Toronto, Canada, May 1991.
6. F. Moscheni, F. Dufaux and M. Kunt, "A new two-stage global/local motion estimation based on a background/foreground segmentation", IEEE Proc. ICASSP'95, vol. IV, pp. 2261-2265, 1995.
7. P. Anandan, J.R. Bergen, K.J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation", Motion analysis and Image Sequence Processing, pp. 1-22. Kluwer Academic Publishers, 1993.
8. T.S.Huang, A.N.Netravali, "Motion and Structure from Feature Correspondences: A Review", Proceedings of IEEE, Vol.82, No.2, February. 1994.
9. M. Bierling, "Displacement estimation by hierarchical block matching", SPIE, Visual Communication and Image Processing'88, vol.1001, pp. 942-951, 1988.
10. C.-T. Chu, "Motion Compensation at VOP Level", M0871, MPEG'96.

Hierarchical Global Motion Estimation/Compensation in Low Bitrate Video Coding

Chung-Tao Chu, Dimitris Anastassiou, Shih-Fu Chang

Image and Advanced Television Laboratory
Department of Electrical Engineering, Columbia University
New York, N.Y. 10027 USA

Abstract

In this paper, we develop a hierarchical global motion estimation/compensation system which is used to improve the block-based motion compensation in the current low bitrate video coding standard H.263. The system first detects the background motion by examining the local motion information using hierarchical block matching and global motion estimation algorithm. If the background motion exists, the global motion compensation (**GMC**) is applied to the image before the H.263 block-based coding. Experiments are done under the typical low bitrate video conditions: low resolution and low frame rate. The results show that prediction errors on both motion vectors and texture are reduced. For sequences containing lots of background motion, the proposed system shows savings of up to 30% in total bitrate.

1. Introduction

High compression ratio video coding has been addressed in ITU-T SG 15/1, which targets video coding for low bitrate communication. It has outlined nearterm specifications for audio-visual coding. The nearterm video coding algorithm H.263 [1] has been optimized in motion prediction and frame formats. Motion estimation is done by block matching Motion vectors are DPCM coded, and motion-compensated predictive coding is used to generate the difference pictures, which are then transform coded using DCT.

One of the features of block-matching motion-compensated video coding at low bitrate is that overhead information usually takes a large portion of its bitstream. Especially, the motion information can take up to 25% of the bitrate, even though the motion vector is also DPCM coded. It is, then, important to reduce the motion information whenever possible.

When background motion exists, it is more likely that motion vector prediction generates errors resulting in more bits used for coding the motion vectors. This happens mostly on the border of the image where the macroblock motion is inconsistent with the other macroblocks inside the image (see figure 1). This issue

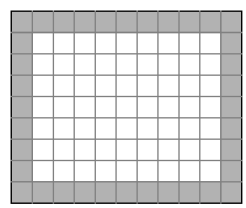


Figure 1: Macroblocks (gray area) **that tend to generate motion vector prediction error.**

becomes more important in low bitrate video coding because the image format is usually smaller, which means that the border macroblocks occupy a higher percentage of the image. Further, the low frame rate, which means large frame difference, results in relatively large background motion and increases the prediction error around the border. The frame distance becomes even

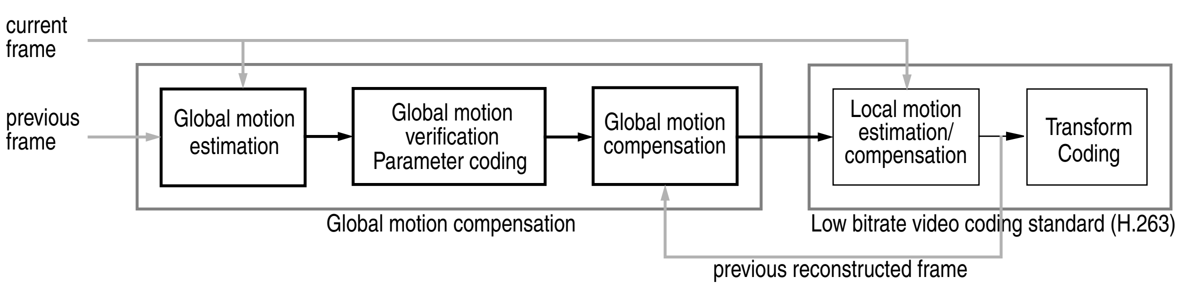
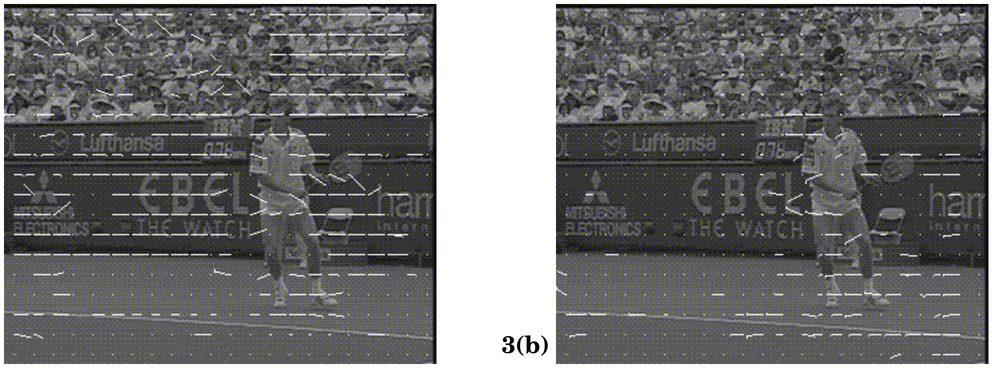


Figure 2: Block diagram of the global motion estimation/compensation.



3(a) 3(b)

Figure 3: Local motion vectors generated by H.263. (pictures from "Stefan" sequence)
3(a) without global motion compensation.
3(b) with global motion compensation.

larger between two P frames if any B-type picture is used.

Inconsistent motion prediction also happens in regions where block matching generates ambiguous results. This is because block matching is optimized to find the minimal block difference instead of true motion. Thus, the local minimum will take over the true motion and contribute to more prediction errors.

Taking into account the above considerations, global motion compensation (see figure 2) can help align the vector field and reduce the prediction error. Figure 3(a) shows the macroblock motion generated by H.263 and figure 3(b) shows the much smoother and more consistent macroblock motions after global motion compensation. Global motion estimation algorithms have been proposed in [2,3,4,5,6]. These algorithms differ in the model to represent the camera motion, as well as in the technique to estimate the parameters of the chosen model. However, in low bitrate video coding with characteristics of low resolution and low frame rate, these models become inefficient to provide good motion prediction, and those algorithms need to be modified for better motion estimation. Here, we choose a more general model that can describe more types of motion. Also we provide an algorithm to estimate the parameters for this motion model [10].

2. Global Motion Compensation

A video sequence can be more efficiently coded if its global motion is compensated first. Even a standard video codec like H.263 can benefit from such technique when there are large and dominant global motions. In [2,4,5] a model of four parameters corresponding to panning and zooming is used. In [3], 2-D rotation is considered and the model uses six parameters. In [6], an affine model with six parameters simplified from an eight parameters model is used. This affine model cannot fully describe 3-D rotation since it uses orthographic projection model and preserves parallelism, and is actually the same as that in [3].

These models are not sufficient when the global motion occurs on a scene with changing depth. Moreover, low bitrate video usually has low frame rate and large frame distance. Large frame distance implies that the global motion can be a combination of several types and the resulting motion may be more complicated. In order to obtain a better prediction on the background motion, a set of eight parameters [7,8] describing the mapping between two images is used to model the global motion (1):

$$S_{f_n}(X, Y) = S_{f_{n-1}}\left(\frac{a_1X + a_2Y + a_3}{a_7X + a_8Y + 1}, \frac{a_4X + a_5Y + a_6}{a_7X + a_8Y + 1}\right) \quad (1)$$

where $S_{fn}(x, y)$ is the pixel value at coordinate (x, y) in frame f_n. This model is capable of describing 3-D translation and 3-D rotation, thus all types of camera motions such as panning, zooming, rotation, and their combinations are included. If background motion exists, the image will be modified using equation (1). Then, the modified image is used as the reference of the H.263 block-based coding.

3. Global Motion Estimation

Given the motion model, the global motion estimation is a problem of least squares minimizations of the motion parameters. There are three techniques for solving this problem. The first one is a differential technique [2,3,7], which uses the information of local gradient and frame difference to find the motion parameters. However, in low bitrate video, it suffers from noises and large motions which happen more often because of larger frame distances. The second one is a matching technique [4,6], which is a generalized block matching that is carried on the background. For a motion model with more parameters, it may be very time consuming in

order to find the best match. The third method is a regression technique [5]. It first estimates the local motion and then uses the local information to find a global motion that minimizes the least square error. Refinements are carried out during the estimation to eliminate noise. In order to reduce the computational load as well as to increase the estimation robustness and accuracy, we propose a hierarchical motion estimation/compensation [10] similar to the third technique, based on the motion model (1). While full-search block matching is widely used to find the minimal block difference, hierarchical block matching, whose result usually approximates true motion, is used in the proposed algorithm. Further, hierarchical motion compensation is used to find the motion parameters with increasing accuracy at each step.

4. Hierarchical Global Motion Estimation/ Compensation

The algorithm is described in figure 4.

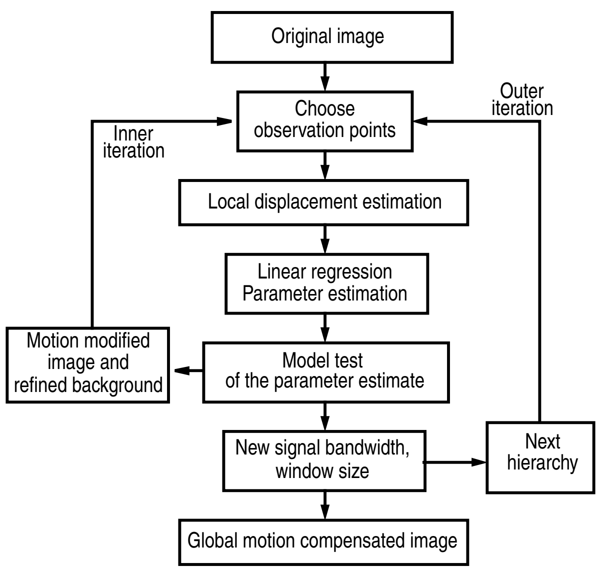


Figure 4. Global motion estimation.

First, equation (1) is re-written into a pair of linear equations. This is a linear system of dimension 8.

$$\begin{cases} X' = a_1X + a_2Y + a_3 - a_7XX' - a_8YX' \\ Y' = a_4X + a_5Y + a_6 - a_7XY' - a_8YY' \end{cases} \quad (2)$$

Second, observation points are chosen throughout the image background and their local motions (displacements from (X, Y) to (X′,Y′)) are estimated. Ideal observation points should be chosen in image areas with higher activity. The local displacements are calculated using hierarchical block matching [10], because it is necessary to estimate the true motion, rather than only minimize displaced frame differences. The estimation uses different measurement window sizes, signal bandwidths, and maximum update displacements on different hierarchical levels. With the hierarchical structure, it is easy to modify the searching range/step for larger motion with few increase in computational load.

Then, the collection of observation points and their local motion information are substituted into the linear equations (2) to solve the motion parameter sets (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8). Four observation points are enough to solve the equations. More observation points and linear regression are introduced to minimize the estimation error. Refinement is carried out during the estimation to eliminate noisy observation points.

After each motion parameter estimation, a model test is performed to modify the image based on the estimated motion. The background region is also refined by exploiting the pixels that do not support the motion estimate. The next iteration of motion estimation is then carried out on the modified image and background region. These inner iterations will continue until no further improvement can be achieved, i.e. a convergent motion is reached.

Hierarchical motion compensation is combined with hierarchical block-matching. That is, for each hierarchy with local motion estimated using different window sizes and signal bandwidths, the global motion is estimated with different accuracy and the motion compensation is carried out. In our experiment, three hierarchies (outer iterations) are used.

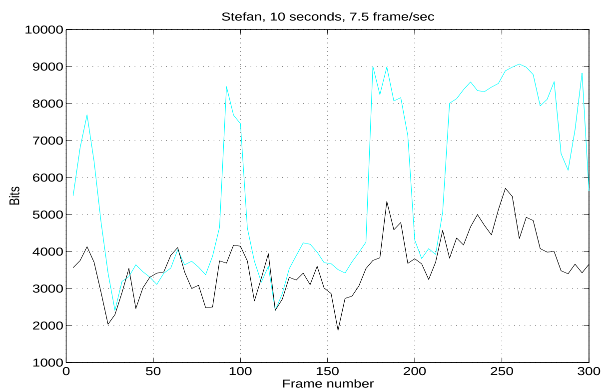
The presented motion estimation features low computing complexity compared with the matching technique. In our experiments, it requires the same order of computational load as those required in H.263 block-based motion estimation. Further, it differs from the mathematical differential model, which solves the linear system using indirect information (local gradient and difference). By using direct information (local displacement), it requires much fewer iterations to converge to a stable solution for the linear system (Eq. (2)).

5. Experimental Results

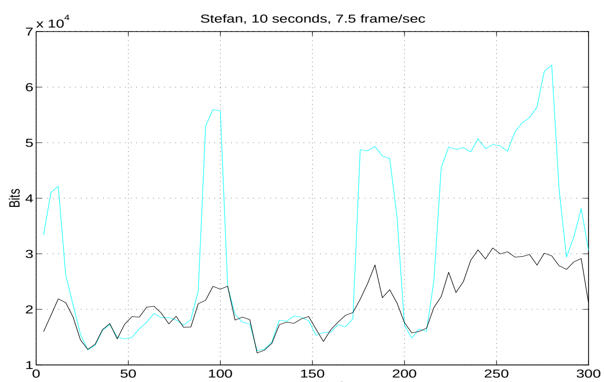
In the real applications, since global motion introduces extra overhead, a verification test is used to evaluate the global motion. Since global motion compensation is more helpful if the motion dominates the image, we use frame difference as the criterion in the verification process. Comparing with the current frame, if the global-motion compensated image has a frame difference of certain threshold less then that of the previ-

ous original frame, the global motion is adopted. Otherwise, only local block-motion compensation is used.

Simulations are performed on "*Stefan*" sequence in CIF format. The international standard H.263 codec is carried out at 7.5 frames/sec with fixed quantization. Experiment results are compared between **H.263** and **GMC+H.263**. Figure 5(a) compares the result of motion information and figure 5(b) compares the total bits coded for each frame. The high peaks of the H.263 curves represent the existence of large background motion. When global motion compensation is used, the motion information can be reduced by as much as 50%. The overall bitrate [**H.263** / **GMC+H.263**] is 233.4/ 163.4 kbps, which is a reduction of 30%.



5(a) comparison of coded motion information



5(b) comparison of total bitrate

Figure 5: Comparison of the total bitrate and vector information coded between H.263: ▬ and GMC+H.263: ▬.

6. Conclusion

A hierarchical global motion estimation/compensation is proposed for low bitrate video coding. The global motion parameters are estimated by using local motion information within the background region. Experiments are done with the standard codec H.263. The results have shown that GMC helps H.263 in reducing motion information as well as overall bitrate when there are large global motions.

References

1. ITU-T SGXV, Expert's Group for Very Low Bitrate Videophone, LBC-95, "Draft Recommendation H.263", July 1995.
2, M.Hotter, "Differential estimation of the global motion parameters zoom and pan", Signal Processing: Image Communication, vol. 16, pp. 249-265, March 1989
3. S.F. Wu and J. Kittler, "A differential method for simultaneous estimation of rotation, change of scale and translation", Signal Processing: Image communication, vol. 2, no. 1, pp. 69-80, May 1990.
4. D. Adolph and R. Buschmann, "1.15 Mbit/s coding of video signals including global motion compensation", Signal Processing: Image communication, vol. 3, nos. 2-3, pp. 259-274, June 1991.
5. Y.T. Tse and R.L. Baker, "Global zoom/pan estimation and compensation for video compression", IEEE Proc. ICASSP'91, vol. IV, pp. 2725-2728, Toronto, Canada, May 1991.
6. F. Moscheni, F. Dufaux and M. Kunt, "A new two-stage global/local motion estimation based on a background/foreground segmentation", IEEE Proc. ICASSP'95, vol. IV, pp. 2261-2265, 1995.
7. P. Anandan, J.R. Bergen, K.J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation", Motion analysis and Image Sequence Processing, pp. 1-22. Kluwer Academic Publishers, 1993.
8. T.S.Huang, A.N.Netravali, "Motion and Structure from Feature Correspondences: A Review", Proceedings of IEEE, Vol.82, No.2, February. 1994.
9. M. Bierling, "Displacement estimation by hierarchical block matching", SPIE, Visual Communication and Image Processing'88, vol.1001, pp. 942-951, 1988.
10. C.-T. Chu, "Motion Compensation at VOP Level", M0871, MPEG'96.