일반논문-07-12-6-08

Global Disparity Compensation for Multi-view Video Coding

Kwan-Jung Oh[a)] and Yo-Sung Ho[a)‡]

ABSTRACT

While single view video coding uses the temporal prediction scheme, multi-view video coding (MVC) applies both temporal and inter-view prediction schemes. Thus, the key problem of MVC is how to reduce the inter-view redundancy efficiently, because various existing video coding schemes have already provided solutions to reduce the temporal correlation. In this paper, we propose a global disparity compensation scheme which increases the inter-view correlation and a new inter-view prediction structure based on the global disparity compensation. By experiment, we demonstrate that the proposed global disparity compensation scheme is less sensitive to change of the search range. In addition, the new inter-view prediction structure achieved about 0.1~0.3dB quality improvement compared to the reference software.

Keyword : MVC, global disparity compensation

Ⅰ. INTRODUCTION

In recent years, various multimedia services have become available and the demand for realistic multimedia systems is growing rapidly. A number of three-dimensional (3D) video technologies, such as two-view stereoscopic system with special glasses, holography, 3D wide screen cinema, and multi-view video have been studied to satisfy these demands.

Among them, MVC (Multi-view Video Coding) is the key technology for various applications including FVV (Free-Viewpoint Video), FTV(Free-viewpoint TeleVision), 3DTV, immersive teleconference, and surveillance. The multi-view video consists of the multi-viewpoint video sequences captured by several cameras at the same time, but different positions. MVC can offers arbitrary viewpoints of dynamic scenes and thus allows more realistic video. Because of the increased number of cameras, the multi-view video contains a large amount of data. Thus, we need to compress the multi-view sequence efficiently without sacrificing visual quality significantly[1-3].

In 2003, the MPEG (Moving Picture Experts Group) Video ad hoc group on 3DAV (3-D Audio and Visual) started works on the MVC standard and now the JVT (joint video team) is working on the MVC standard[4].

The key problem in multi-view video coding is how to reduce the inter-view correlation efficiently. Inter-view correlation which is correlations between frames captured at the same time in different cameras is quite different from the temporal correlation which is correlation between frames captured in the same camera at different times. Actually, while traditional ME(motion estimation)/MC(motion compensation) is a good model for predicting temporally adjacent frames, it is less accurate for an inter-view prediction. Because of the global disparity, previous MVC

a) 광주과학기술원 정보통신공학과
Gwangju Institute of Science and Technology (GIST) Department of Information and Communications
‡ 교신저자 : 호요성(hoyo@gist.ac.kr)
※ 본 연구는 광주과학기술원(GIST) 실감방송연구센터(RBRC)를 통한 정보통신부대학IT연구센터(ITRC)와 교육인적자원부 두뇌한국21(BK21) 정보기술사업단의 지원에 의한 것입니다.

scheme used a fairly large search range for ME process in not only inter-view prediction but also temporal prediction. It causes an inefficient coding and much encoding time.

In this paper, to solve these problems we propose a global disparity compensation for multi-view video coding and a new inter-view prediction structure using global disparity compensation. We previously calculate the global disparity between certain two views considering inter-view prediction structure. And then, before the ME/MC we shift reference frames as much as its global disparity. After the shifting of reference frames, we fill empty pixels caused by shifting.

Ⅱ. MULTI-VIEW VIDEO CODING (MVC)

1. General Multi-view Video System

Multi-view video system contains the process from the acquisition to the display of multiple video sequences. Figure 1 shows the general multi-view video system.

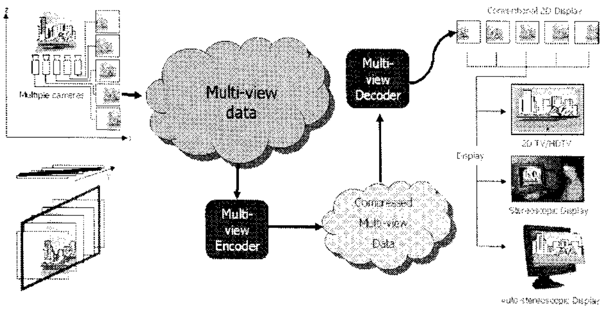


그림 1. 일반적인 다시점 비디오 시스템
Fig. 1. General Multi-view Video System

At first, we acquire multi-views sequences by using multiple cameras. And then, MVC encoder compresses the multi-view video data. The encoded bitstream is transmitted through the channel. The MVC decoder converts encoded bitstream to multi-view video sequences. Finally, one display device is chosen by its application among the several devices.

2. Reference Coding Scheme

Proposals received in response to the CfP(Call for Proposals) have shown that specific MVC technology outperforms the AVC reference solution(simulcast anchors used in CfP) significantly in terms of PSNR and subjective quality. Different view-temporal prediction structures as well as specific MVC tools have been proposed that are promising for inclusion in a future MVC standard. Fraunhofer-HHI suggested the MVC scheme based on JSVM (Joint Scalable Video Model) and it is adapted reference scheme for MVC standardization.

The reference coding scheme, as shown in Fig. 2, uses a prediction structure with hierarchical B pictures for each view. Additionally, inter-view prediction is applied to every 2nd view, i.e. S1, S3 and S5 in Fig. 2. For an even number of views, the prediction scheme of the last view (S7 in Fig. 1) is a mix of even and odd views. As there is just one neighboring view for P view and B view have two inter-view references. Inter-view prediction structure is starts and ends with P view.

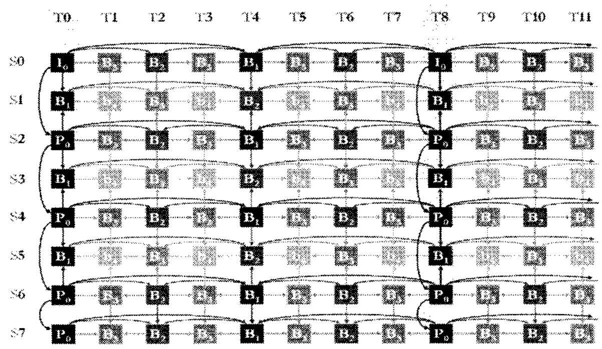


그림 2. 참조 부호화 모델의 시·공간적 예측 구조
Fig. 2. View-temporal Prediction of Reference Scheme

Ⅲ. PROPOSED ALGORITHMS

1. Global Disparity

Multi-view video consists of several videos captured by multiple cameras. So, there exists a global disparity between adjacent views.

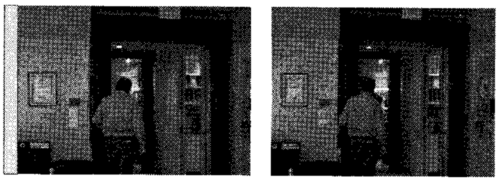


그림 3. "Exit" 시점 0과 시점 1의 전역 변이
Fig. 3. Global Disparity for "Exit" view 0 and view 1

Figure 3 shows the global disparity between "Exit" view 0 and view 1. "Exit" view 1 looks like the shifted version of "Exit" view 0 as much as the shaded area.

2. Global Disparity Calculation

To calculate the global disparity, we employ MAD (Mean Absolute Difference) in Eq. (1) and Fig. 4 shows related parameters and overlapped images. The minimum (g_x, g_y) is the global disparity.

$$(g_x, g_y)_{MAD} = \min_{x,y}\left[\frac{1}{R}\sum_{i,j\in R}\left|img0(i,j) - img1(i-x, j-y)\right|\right] \qquad (1)$$

img0 and *img1* in Fig. 4 are two pictures for global disparity calculation and R is the number of pixels in the overlapped area.

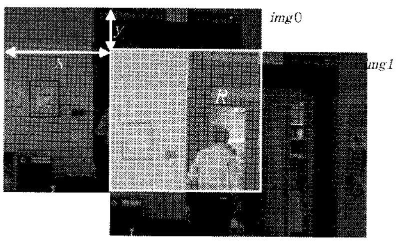


그림 4. 전역 변이 계산
Fig. 4. Global Disparity Calculation

3. Global Disparity Compensation

By using the calculated global disparity value, we compensate the global disparity. Before motion estimation process, we shift the reference frame as much as the global disparity. And then, we pad empty outside of the boundary by copying the boundary pixel value. From the Fig. 5 to Fig. 9, you can easily understand the procedure of global disparity compensation. Figure 5 is the frame to be encoded as B frame, and Fig. 6 shows the reference frames for that. There exist the global disparities between pictures in Fig. 5 and Fig. 6. So, we need large search range in the motion estimation process to search the proper region in general.

그림 5. B화면으로 부호화 되어야 할 화면
Fig. 5. Frame to be encoded as B frame

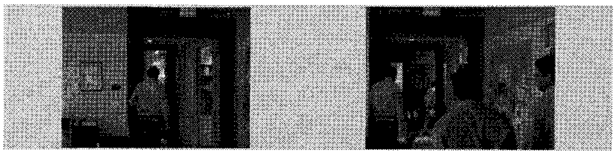


그림 6. 그림 5의 두 참조화면
Fig. 6. Two Reference Frames for Fig. 5

However, if we shift the reference frames as much as the global disparity, we do not need to a large such range anymore. Figure 7 shows the shifted reference frames. Most objects in the pictures are located at similar positions compared to the Fig. 7 but some regions are empty. We pad the empty regions by copying the boundary values as shown in Fig. 8 or by copying from the other reference frame as shows in Fig. 9.

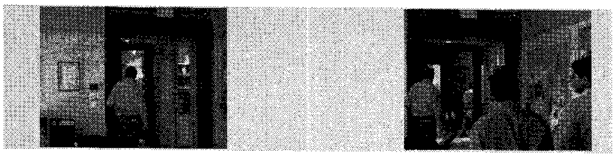


그림 7. 이동된 참조 화면들
Fig. 7. Shifted Reference Frames

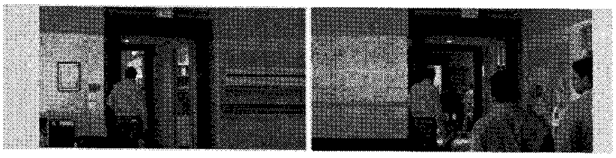


그림 8. 가장자리 영역을 복사하여 빈 영역 채움
Fig. 8. Filling the Empty Region by Copying the Boundary Pixel Value

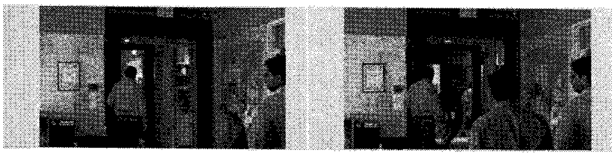


그림 9. 다른 참조화면을 이용하여 빈 영역 채움
Fig. 9. Filling the Empty Region by Using Other Reference Frame

4. A New Inter-view Prediction Structure

In general, inter-view prediction structure can not uses the far distance prediction because their coding efficiencies are not good. However, we propose a new inter-view prediction structure including a far distance prediction by using global disparity compensation. We apply the hierarchical-B picture structure to the inter-view prediction by using global disparity compensation[6]. Figure 10 shows the proposed inter-view prediction structure using global disparity compensation in case of 8 views sequence.

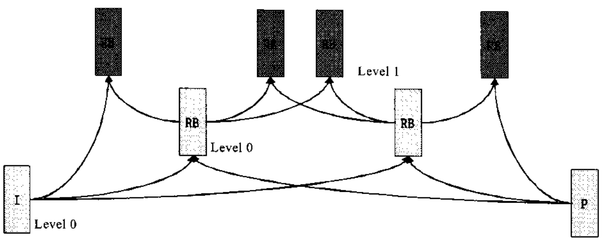


그림 10. 새로운 시점 부호화 구조
Fig. 10. A New Inter-view Prediction Structure

Ⅳ. EXPERIMENTAL RESULTS

Following simulation demonstrates the existence of the global disparity. We analyze the coding results for various search range. We use the IPPP GOP structure, QP=31, and view-interlaced structure as shown in Fig. 11.

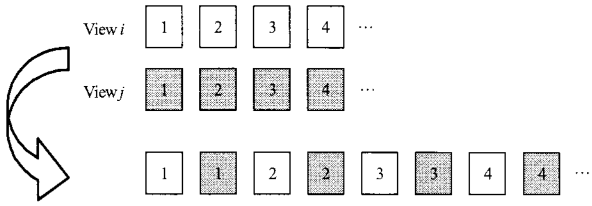


그림 11. 시점-교차 구조
Fig. 11. View-interlaced Sequence

As shown in Table 1, coding efficiency increases up to a certain search range, however it is saturated when search range is over the certain search range. Because certain search range is over the current global disparity and it is an evidence of global disparity.

표 1. "Exit" 시점 3과 시점 5에 대한 부호화 결과
Table 1. Coding Results for "Exit" view 3 and view 5

Search Range	PSNR (dB)	Bitrate (kbps)
16	38.05	1340
32	37.92	1171
64	37.88	1114
96	37.88	1111
128	37.87	1105
160	37.88	1108

To demonstrate the efficiency of the global disparity compensation, we introduce one simple experiment and show its results. We use the full search mode in motion estimation instead of the fast search mode and encode just three frames using I-B-I structure for "Exit" sequence. A quantization parameter (QP) is 31.

Table 2 shows the coding results of B picture when search range varies from 32 to 128. In this case, global disparities are (-75, 0) and (126, 0). As you can see, the proposed algorithm shows better results and it is less sensitive to change of the search range. Therefore, we can know that global disparity compensation improves the coding efficiency of MVC and it also leads to shorter encoding time. The global disparities are transmitted in the MVC bit stream for recreating the reference frames at decoder side. Table 3 shows the encoding time comparison for various search ranges.

표 2. 전역 변이 보상에 대한 부호화 결과
Table 2. Coding Result for Global Disparity Compensation

Search Range	Bit Rate (bits)		PSNR (dB)	
	JSVM	Proposed	JSVM	Proposed
32	56272	45096	38.1103	38.0668
64	49096	44224	37.9909	38.0532
96	45480	44360	37.9919	38.0622
128	44136	44352	38.0649	38.0589

표 3. 다양한 탐색 범위에 대한 부호화 시간 비교
Table 3. Encoding Time Comparison for Various Search Ranges

Search Range	32	64	96	128
Encoding Time (sec.)	415	1643	3624	6537

In order to evaluate efficiency of the new inter-view prediction structure, we have experimented on only inter-view frames with "Ballroom" and "Race1" sequences and compared to the result of the JSVM coding. Table 4 shows the coding results and Fig. 12 shows the rate-distortion curve for the "Ballroom" sequence. In this paper, we ignore the bits for global disparities coding.

표 4. "Ballroom"에 대한 부호화 결과
Table 4. Coding Result for "Ballroom"

Basis QP	Bit Rate (kbps)		PSNR (dB)	
	JSVM	Proposed	JSVM	Proposed
31	1001.075	908.975	37.2452	37.0375
29	1290.175	1160.000	38.0992	37.8589
26	2031.975	1804.925	39.4265	39.1562

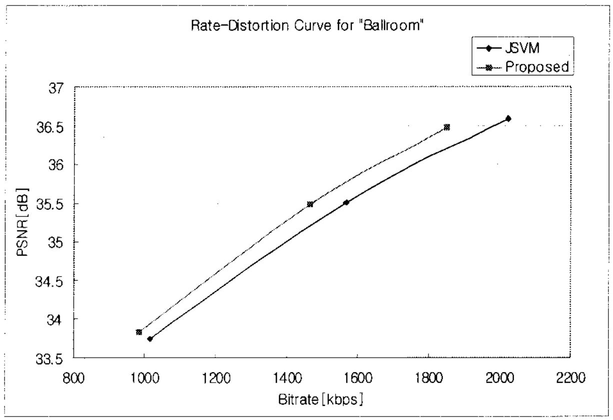


그림 12. "Ballroom"에 대한 비트율-왜곡 곡선
Fig. 12. Rate-Distortion Curve for "Ballroom"

표 5. "Race1"에 대한 부호화 결과
Table 5. Coding Result for "Race1"

Basis QP	Bit Rate (kbps)		PSNR (dB)	
	JSVM	Proposed	JSVM	Proposed
28	1722.84	1578.00	39.6925	39.5550
26	2217.93	2002.83	40.8414	40.6218
24	2953.44	2641.53	42.0788	41.8026

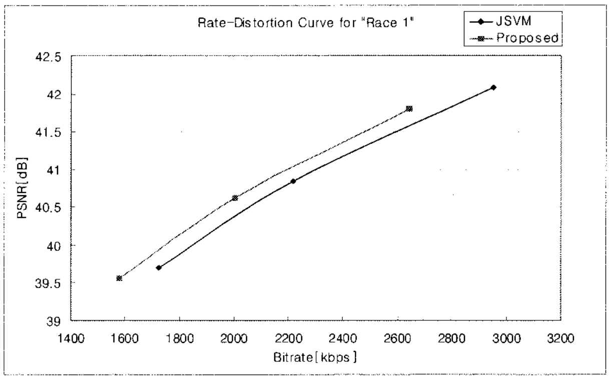


그림 13. "Race1"에 대한 비트율-왜곡 곡선
Fig. 13. Rate-Distortion Curve for "Race1"

Table 5 and Fig. 13 show the coding results and the rate-distortion curve for "Race1" sequence, respective- ly. As you can see, the proposed algorithm is much better than the reference coding scheme. The proposed algorithm achieved about 0.1~0.3 dB quality improvement compared to the reference coding scheme (JSVM).

V. CONCLUSIONS

In this paper, we have proposed the global disparity compensation for multi-view video coding and the new inter-view prediction structure based on global disparity compensation. With some experiments, we have verified improvement of the proposed algorithms. The smaller search range is used, the better coding results are shown. The new inter-view prediction structure achieved about 0.1~0.3dB quality improvement compared to the reference software.

REFERENCES

[1] A. Smolic and P. Kauff, "Interactive 3D Video Representation and Coding Technologies,"Proceedings of the IEEE, Spatial Issue on Advances in Video Coding and Delivery. Vol. 93. pp. 99-110, 2005.

[2] A. Smolic, K. Mueller, T. Rein, P. Eisert, and T. Wiegand, "Free Viewpoint Video Extraction, Representation, Coding, and Rendering," Proceeding of IEEE International Conference on Image Processing. Vol. 5. pp. 3287-3290, 2004.

[3] R. S. Wand and Y. Wang, "Multiview Video Sequence Analysis, Compression, and Virtual Viewpoint Synthesis," IEEE Transactions on Circuit and System for Video Technology, Vol. 10. pp. 397-410, 2000.

[4] ISO/IEC JTC1/SC29/WG11 W8019, "Description of Core Experiments in MVC," April 2006.

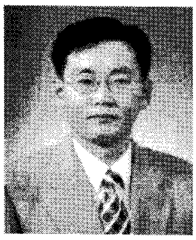
[5] ISO/IEC JTC1/SC29/WG11 n6501, "Requirements on multi-view video coding," October 2004.

[6] ISO/IEC JTC1/SC29/WG11 M12542, "Multi-view Video Coding based on Lattice-like Pyramid GOP Structure," October 2005.

저 자 소 개

Kwan-Jung Oh

- 2002년 8월 : 전남대학교 정보통신공학부 (학사)
- 2005년 2월 : 광주과학기술원 정보통신공학과 (석사)
- 2005년 3월 ~ 현재 : 광주과학기술원 정보통신공학과 박사과정
- 주관심분야 : 다시점 비디오 부호화(MVC), 자유시점 TV(FTV), H.264/AVC, 3차원 TV, 영상신호 처리 및 압축, 실감방송



Yo-Sung Ho

- 1981년 2월 : 서울대학교 전자공학과 (학사)
- 1983년 2월 : 서울대학교 전자공학과 (석사)
- 1983년 3월 ~ 1995년 9월 : 한국전자통신연구원 선임연구원
- 1989년 12월 : University of California, Santa Barbara, Department of Electrical and Computer Engineering (박사)
- 1990년 1월 ~ 1993년 5월 : 미국 Philips 연구소 Senior Research Member
- 1995년 9월 ~ 현재 : 광주과학기술원 정보통신공학과 교수
- 주관심분야 : 디지털 신호처리, 영상신호처리 및 압축, 멀티미디어 시스템, 디지털 TV, MPEG 표준, 3차원 TV, 실감방송