

自由視点映像生成に適した N-View N-Depth 符号化

樋口 雄太[†] メヒルダドパナヒプル テヘラニ[†] 圓道 知博^{††} 藤井 俊彰[†] 谷本 正幸[†]

[†] 名古屋大学 工学研究科

^{††} 長岡技術科学大学 工学科 電気系

E-mail: [†]higuchi@tanimoto.nuee.nagoya-u.ac.jp, ^{††}{panahpour,fujii,tanimoto}@nuee.nagoya-u.ac.jp,
^{†††}yendo@nagaokaut.ac.jp

あらまし 本報告では、多視点映像及び各視点に対応する奥行き情報の圧縮符号化について取り扱う。復号後に行う自由視点映像生成に適した符号化方式を提案する初期段階として、多視点映像及び奥行き情報各々の伝送レートが自由視点映像生成の精度に及ぼす影響について検討する。現在標準化団体 MPEG において、復号側での自由視点映像生成を前提とした、新たな多視点映像及び奥行き情報符号化方式が議論されている。この方式では、全体のデータレートがある値に定められた際、その範囲中で多視点映像及び奥行き情報各々のビットレートを決定しなくてはならない。その際、伝送量が復号後に行われる自由視点映像生成の精度が最大となる様、各ビットレートを適した値に調整し、伝送する必要がある。実験では、多視点映像及び奥行き情報の符号化方式として Multiview Video Coding (MVC) を並列に使い、復号情報より生成された視点映像と、同位置で撮影された映像との比較を行った。結果より、特に総伝送量が一定である場合、多視点映像に比べ奥行き情報は低ビットレートでも視点映像生成の精度に影響を及ぼしにくい事が分かった。

キーワード 多視点動画像符号化, 奥行き情報符号化, 自由視点映像生成

N-View N-depth Coding for Free View Generation

Yuta HIGUCHI[†], Mehrdad PANAHPOUR TEHRANI[†], Tomohiro YENDO^{††}, Toshiaki FUJII[†],
and Masayuki TANIMOTO[†]

[†] Graduate School of Engineering, Nagoya University

^{††} Department of Electrical Engineering, Nagaoka University of Technology

E-mail: [†]higuchi@tanimoto.nuee.nagoya-u.ac.jp, ^{††}{panahpour,fujii,tanimoto}@nuee.nagoya-u.ac.jp,
^{†††}yendo@nagaokaut.ac.jp

Abstract In this paper, we discuss a multiview video and depth coding for Multiview applications such as 3DTV and Free View-point Television. We investigate the effect on free view synthesis quality by changing the transmission rates between multiview and depth sequences. In the simulations, we employ MVC in parallel to compress the multiview video and depth sequences at different bitrates, and compare the virtual view sequences generated by decoded stream with the original video sequences taken in the same viewpoint. Our results show that multiview depth bitrates has less effect on the view synthesis quality compared with the multiview video bitrates.

Key words Multiview Video Coding, Depth Coding, Free View Generation

1. まえがき

近年、多視点映像を用いたシステム及びアプリケーションに関する研究が多数扱われている。これらのアプリケーションは、複数台のカメラで撮影した映像を使用し、従来の映像コンテンツと比べより没入感のある体感型コンテンツを視聴者に対し提供できる。本研究室にて提案している Free viewpoint

Television (FTV) [1] は、多視点映像及び各々の視点に対応する奥行き情報より、実際には撮影されていない仮想視点の映像を生成する。この技術により、視聴者は個人が望む視点を自由に選択する事ができ、より自由度の高いコンテンツを楽しむ事が可能となる。しかし、多視点映像システムを実現する為には、カメラキャリブレーション、奥行き情報推定、多視点映像データの伝送等、様々な技術課題が存在する。その中でも多視点映

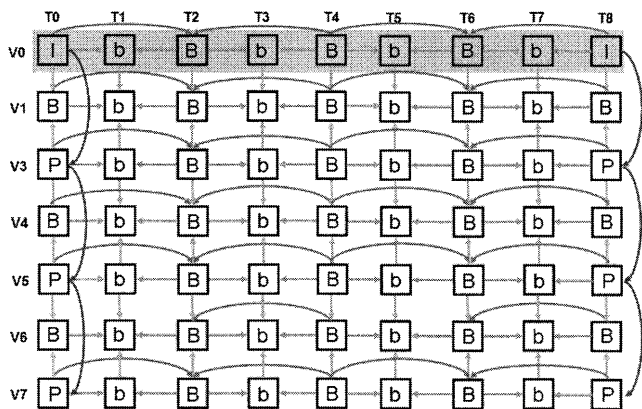


図 1 Multiview coding structure (Num of view = 8, GoP = 8)

像データの圧縮伝送は、多視点映像システムの実現に重要な技術の一つである。多視点映像は撮影した視点数分の映像から構成される為、単視点映像と比べデータ量が膨大となる。よって、多視点映像コンテンツを視聴者に対し配信するには、多視点映像システムに適した、より圧縮性能の高い符号化方式が必要不可欠である。

多視点映像符号化技術として、Joint Video Team (JVT) より Multiview Video Coding (MVC) [2] が標準化されている。MVC は多視点映像のみを入出力とする方式であるため、復号側で多視点映像システムを提供するには、各視点に対応する奥行き情報を別途取得する必要がある。そこで現在標準化団体 MPEG より、新たな多視点映像符号化技術として、多視点映像及び奥行き情報を入出力とする符号化方式の標準化が行われている。この符号化方式では、復号側にて多視点映像システムが用いられる事を想定しており、復号されたデータから自由視点映像生成を行うことが考えられている。しかし、多視点映像及び奥行き情報両方のシーケンスを送送することで、さらなる伝送量の肥大化が生じてしまう。また、総伝送レートがある値に定められた際、その範囲内で多視点映像及び奥行き情報各々のビットレートを決定しなくてはならない。その際、伝送量が復号後に行われる自由視点映像生成の精度が最大となる様、各シーケンスのビットレートを符号器で適した値に設定し、伝送する必要がある。

本報告では、復号後の自由視点映像生成に適した符号化方式について検討する初期段階として、多視点映像及び奥行き情報各々の伝送レートが自由視点映像生成に及ぼす影響について考察する。実験では、多視点映像及び奥行き情報の符号化方式として Multiview Video Coding (MVC) を並列に用い、復号情報より生成された視点映像と、同位置で撮影された映像との比較を行った。

本稿の構成は以下の通りである。まず 2 章にて Multiview Video Coding の概要について説明する。3 章では、仮想視点映像生成に適した奥行き情報の符号化方式について述べ、本報告にて取り扱う N-View N-Depth 形式の圧縮符号化の概要を説明する。4 章で実験とその結果を示し、最後に 5 章で結論を述べる。

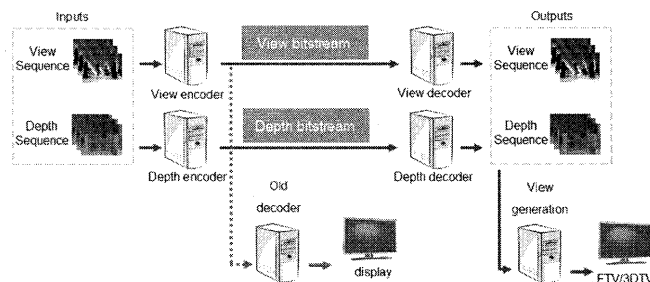


図 2 Diagram of N-View N-Depth coding system which we consider

2. Multiview Video Coding

3DTV や FTV 等の多視点映像システムは、カメラアレイ等で撮影された複数の視点映像及びカメラ位置情報等を入力とする。これらの入力は従来の単視点映像と比べ膨大なデータ量を持つため、視聴者側にこれらの情報を伝送する為にはより効率の良い符号化方式が必要である。多視点映像は同シーンを別視点から撮影した映像から構成されており、時間方向だけでなく視点方向に対し強い相関性を持つ。Multiview Video Coding (MVC) [2] は多視点映像が持つ視点方向への相関を利用し、従来の動画符号化方式である H.264/AVC [3] と比べて、より高い圧縮性能を実現している。MVC の符号化構造は、H.264/AVC の符号化構造に視差補間予測を新たに採用した形で表される。ここで視差補償予測とは、既に符号化済みの視点映像から符号化対象としている視点映像を比較し、より相関度の高い比較映像と符号化対象映像との差分を符号化するものである。隣り合う視点の映像間に生じる視差は、カメラ間隔が十分近ければ密な動き空間とみなせる。よって、時間方向への相関除去に利用される動き補償予測アルゴリズムを、視点方向のフレーム間予測に適用することで、視点方向に生じる相関を除去できる。

MVC において、符号化する視点数を 7,GoP サイズを 8 とした場合の符号化構造を図 1 に示す。ここで、 T_n 及び V_n はそれぞれ各フレームが属するタイムスタンプ及び視点位置を示している。また、横方向に伸びる矢印は時間方向へのフレーム間予測、横方向に伸びる矢印は視点方向へのフレーム間予測を表している。MVC は H.264/AVC を拡張した形である為、MVC の符号化構造では H.264/AVC と類似したピクチャ構造、I ピクチャ(I)、P ピクチャ(P)、B ピクチャ(B)が採用されている。I ピクチャでは符号化対象ピクチャ自身から予測され、P 及び B ピクチャは異なる時間・視点に属するピクチャから予測される。図の符号化構造において視点 V_0 だけに注目すると(図 1 内の斜線部)、H.264/AVC と同じ符号化構造をしていることが分かる。これらの技術により、MVC は H.264/AVC を各視点映像に対し並列に用いた場合より 20~30% の符号化率向上を実現している。

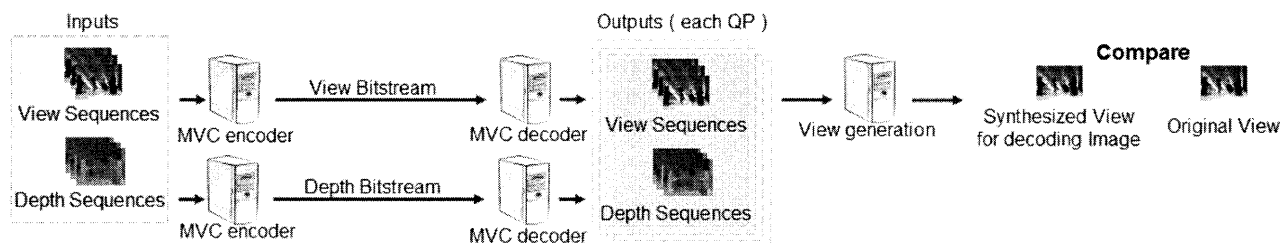


図 3 Diagram of experimental environment

3. 自由視点合成に適した N-View N-depth 符号化方式

MVCにより符号/復号化されたデータのみ使用するだけでは、依然として視聴者は事前に撮影された視点映像のみしか選択する事が出来ない。これは、撮影コストやカメラの物理的な大きさにより、多視点映像がより密に取得できない為である。これらの理由により、視聴者側で視点を自由に選択可能とする為には、多視点映像システム内で、復号された情報を用いて事前に撮影されていない仮想視点の映像を生成する必要がある。自由視点映像生成には、多視点映像に加えて各々の視点に対応する奥行き情報を必要とする。奥行き情報は、カメラと映像内のオブジェクトとの距離を一般的な 8bit グレースケール映像として表現したものである。MVCを用いて多視点映像を伝送する場合、復号側では視点映像のみしか受け取れない。よって、復号側にて自由視点映像生成を行うには、何らかの手法を用いて別途奥行き情報を取得する必要がある。奥行き情報は復号された視点映像より推定できるが、計算負荷が非常に高いため、コンテンツ視聴の際に遅延等の原因となる。復号側で多視点映像システムをより低負荷で実現する為、本報告では多視点映像及び奥行き情報の両方を符号化し、伝送する事を考える。これにより復号側では、視聴者が奥行き推定無しでより滑らかな自由視点映像コンテンツを提供する事ができる。

本報告で扱うシステム全体の概略を図 2 に示す。符号器は、撮影された多視点映像及び奥行き情報を入力とする。ここで入力情報は、事前に幾何補正・色補正が行われているものとする。復号器は、入力と同じ視点数を持つ、多視点映像及び奥行き情報を出力する。復号後、復号データを基に自由視点映像生成を行い、視聴者の要求に応じて仮想視点映像を生成する。本報告では、MVCを用いて多視点映像及び奥行き情報各々の符号化を行う。多視点映像を MVCを用いて符号化することにより、H.264/AVC等以前に標準化された復号器に対する互換性を持たせる事が出来る。奥行き情報はグレースケール映像として表現されているため、多視点映像の符号化とは別に、MVCを用いて独立に符号化できる。しかし、奥行き情報は自由視点映像生成の精度に対し強い影響を持つため、自由視点映像生成に適した形で符号化しなくてはならない。加えて、符号化情報の総伝送レートがある一定の値で与えられた場合、その中で多視点映像及び奥行き情報各々のデータを送る割合を調整する必要がある。その時、自由視点映像生成の精度が最大となるよう、

各データを適したビットレートで符号化しなくてはならない。

4. 実験

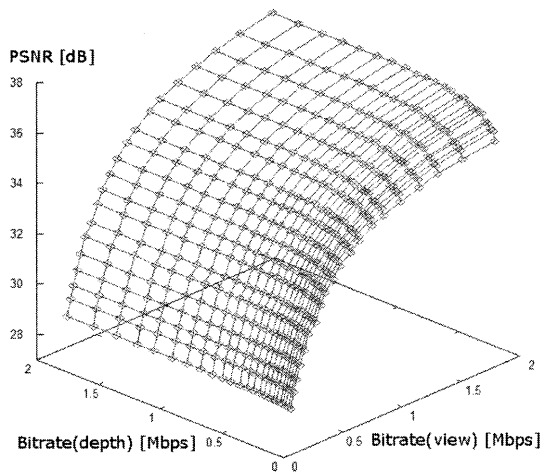
前章で述べた符号化方式において、多視点映像及び奥行き情報各々の伝送レートが自由視点映像生成に及ぼす影響について考察する。行った実験の概略を図 3 に示す。多視点映像及び奥行き情報を圧縮率を変えて符号化し、復号データより生成した仮想視点映像の精度を評価した。以下の流れで実験を行った。

- (1) 多視点映像及び奥行き情報を MVCを用いて並列に符号化する。この時、視点映像及び奥行き情報各々の量子化パラメータを 30~51 の間で変化させ、異なるビットレートにて圧縮する。
- (2) 複数得られた復号データより仮想視点映像を生成
- (3) 生成した仮想視点映像と、同じ位置で実際に撮影された映像とを比較。

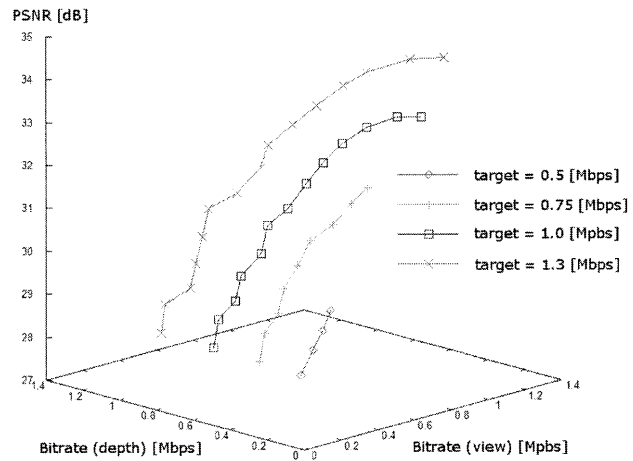
MVC アルゴリズムとして、JVT より提供されているリファレンスソフトウェア JMVC8.3 を使用した。設定した符号化パラメータを以下に示す。

- gopsize : 15
- intraperiod : 15
- searchmode : fastmode
- searchfuncfullpel : SAD-YUV
- searchfuncsubpel : hadamard
- searchrange : 32

上記で明示していない項目は全て初期値に設定した。符号化するシーケンスとして、HHI より提供されている多視点映像シーケンス "BookArrival" (1024 × 768 pixel) から視点 6, 8, 10 の 3 視点を用いた。各々の視点は直線上に 10cm 間隔で配置されており、幾何補正が済んでいるものとする。自由視点映像生成では、復号データを用い、視点 6 と 8 より視点 7 を、視点 8 と 10 より視点 9 を生成する。自由視点映像生成のアルゴリズムとして、MPEG より提供されているリファレンスソフトウェア VSRS3.5 [4] を使用した。VSRS3.5 は、生成したい視点映像の左及び右に位置する視点映像、それぞれの視点に対応する奥行き情報、カメラパラメータを入力とし、指定された位置の視点映像を出力する。VSRS3.5 で設定したパラメータを以下に示



(a) Result in "BookArrival" of view 7



(b) Result in "BookArrival" of view 7 in each target

図 4 Bitrate (view/depth) versus PSNR

表 1 The PSNR peak at each target rate

Target[Mbps]	PSNR[dB]	bitrate(v)[Mbps]	bitrate(d)[Mbps]
0.5	29.6577	0.3210	0.1742
0.75	32.2503	0.5368	0.1894
1.0	33.6108	0.7482	0.2363
1.3	34.6416	1.0231	0.2572

表 2 PSNR values and each bitrates (target = 1.3[Mbps])

Label	PSNR[dB]	bitrate(v)[Mbps]	bitrate(d)[Mbps]
Combination 1	34.64	1.0231	0.2572
Combination 2	34.00	0.7482	0.5308
Combination 3	32.15	0.5368	0.7294

す。

- Precision : Bi-cubic
- intraperiod : Not Blend

上記で明示していない項目は全て初期値に設定した。

図 4 は、多視点映像及び奥行き情報各々のビットレート対 PSNR のグラフである。ここで、図 4 における視点映像及び奥行き情報のビットレートは、それぞれのシーケンスにおける 3 視点分の合計ビットレートである。伝送可能な総ビットレートを定められている場合を考える際は、目標の総ビットレートを *target* とすると、

$$target = Bitrate(view) + Bitrate(depth)$$

の式で表される平面とグラフ表面の交線に属するビットレートの組み合わせにおいて、一番 PSNR の高い候補を選択することが望ましい。

初めに、多視点映像及び奥行き情報個々のビットレート変化が自由視点映像生成に及ぼす影響について検討する。図 4(a) に "BookArrival" 視点 7 における、各シーケンスの圧縮レート対生成した視点と撮影された視点映像との PSNR 特性を示す。視点 7 の映像は、復号された視点 6 及び視点 8 の映像及び各々の視点に対応する奥行き情報より生成したものである。図 4(a) より、自由視点映像生成の精度が殆ど視点映像の符号化精度に依存していることが見て取れる。奥行き情報のビットレート変動に着目すると、奥行き情報の総ビットレートが 0.6[Mbps] を下回る付近まで生成精度に変化が見られない。また、MVC に

おける量子化パラメータの最大値は 51 であり、仕様上ではこれ以上ビットレートを下げることができない。これらの事より、多視点映像に比べ、奥行き情報は低ビットレート時でも自由視点映像生成の精度に影響しにくい事が分かる。

次に、伝送レートが一意に定められた場合について検討する。図 4(b) に示す各直線は、図 4(a) 中で、それぞれ定められた総伝送レートを満足する多視点映像及び奥行き情報のビットレートを組み合わせて生成した候補を表している。また表 1 に、図 4(b) 中の総伝送レートにおいて、一番 PSNR が高い候補を示す。これらの結果より、図中のどの伝送レートにおいても、奥行き情報が全体に占める割合が非常に低い候補で、一番自由視点映像生成の精度が高くなっている事が分かる。ここで、生成した視点 7 の映像の一部と、生成の際に使用した視点 6 に対応する奥行き情報を図 5 に示す。また、図 5 に示す候補における生成結果の PSNR 値と、使用した多視点映像及び奥行き情報各々のビットレートを表 2 に示す。図 5 及び表 2 において、Combination 1 は伝送レートが 1.3[Mbps] の時に一番 PSNR が高い候補である。Combination 2 及び 3 は、Combination 1 より奥行き情報のビットレートが高い組み合わせで生成された候補である。図 5(a) に示す生成映像結果を見ると、オブジェクト境界部分での崩壊が顕著に現れている。これらの合成エラーが生じる原因として、視点映像生成の際に使用する奥行き情報が非常に粗いものである事が挙げられる。実際、図 5(a) に示す奥行き情報を見ると、高い圧縮による歪が生じている事が見て取れる。それに対し図 5(b) 及び 5(c) を見ると、合成に使用している奥行き情報は、PSNR が一番高い候補に比べてより歪が少ない事が分かる。従って、これらの候補によって生成さ

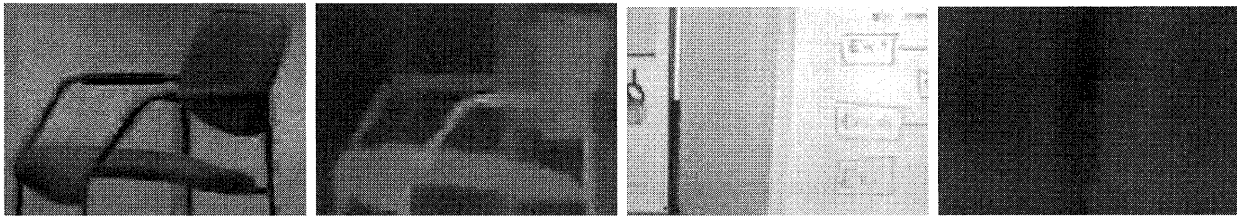
れた仮想視点映像は、オブジェクト境界部分がより正確に表現されている。しかし、自由視点映像生成に使用された多視点映像のビットレートが低くなっているため、生成された映像、特にテクスチャが複雑な部分に大きな歪が生じてしまっている。

5. ま と め

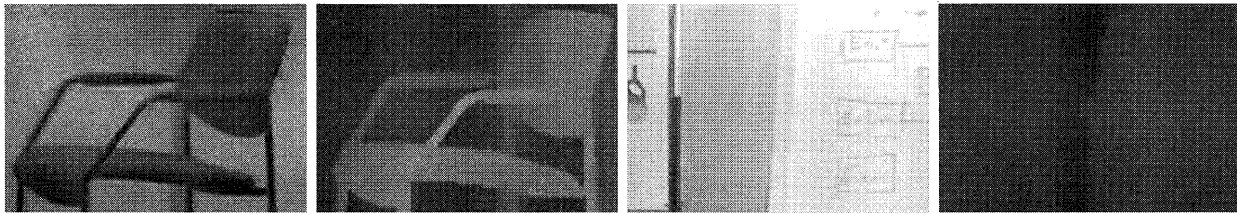
本報告では、複数の視点映像及び各々の視点に対応する奥行き情報を入出力とする N-view Ndepth 圧縮符号化において、多視点映像及び奥行き情報各ビットレートの違いが及ぼす自由視点映像生成への影響について考察した。実験では、多視点映像及び奥行き情報を MVC を用いて独立に符号化し、復号情報より生成した仮想視点映像と実際に撮影された映像との比較を行った。実験結果より、奥行き情報は多視点映像に比べて低ビットレートで伝送されても、自由視点映像生成の精度に影響を与えにくいことが分かった。特に伝送レートが一意に与えられた場合、奥行き情報が占める割合をより少なくした際に、より高い PSNR 値を示す事が分かった。今後の課題として、多視点映像及び奥行き情報の間に存在する相関を利用した、より適した符号化方式の提案が挙げられる。本報告では、多視点映像及び奥行き情報を独立に符号化したが、双方の情報には依然として類似性が残ったままである。よって、これらの相関性を利用することで、より圧縮効率の向上が見込めるのではないかと考える。

文 献

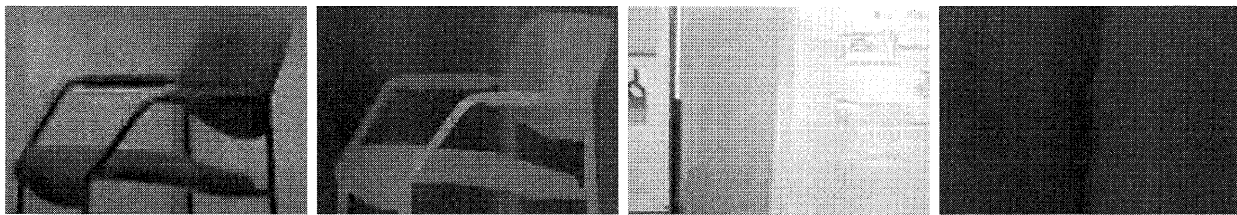
- [1] T.Fujii and M.tanimoto, "Free-viewpoint Television based on the Ray-Space representation", SPIE ITCOM'02, pp. 175-189, Aug.2002.
- [2] Y.Chen, M.M.Hannuksela, L.Zhu, A.Hallapuro, M.Gabbouj and H.Li, "Coding Techniques in Multiview Video Coding and Joint Multiview Video Model", PCS2009, May.2009.
- [3] Thomas Wiegand, Gary J. Sullivan, Gisle Bjontegaard, and Ajay Luthra, "Overview of the H.264/AVC Video Coding Standard", IEEE Trans. Circuits & Syst. for Video Technology, 13, No. 7, pp.560-576, Jul.2003.
- [4] M.Tanimoto, T. Fujii, and K. Suzuki, "View Synthesis Algorithm in View Synthesis Reference Software 2.0 (VSRS2.0)", ISO/IEC JTC1/SC29/WG11 M16090, Lausanne, Switzerland, February 2008.



(a) Combination 1



(b) Combination 2



(c) Combination 3

図 5 Synthesis result and quality of used depth sequences. (*target* = 1.3[Mbps])