

## 視差情報を用いた多眼 3 次元画像の効率的符号化法

八木原英之      浜村 倫行      相田 仁      齊藤忠夫

東京大学大学院工学系研究科

ステレオ画像の符号化方式として、MPEG2 の時間スケーラビリティ機能を用いたマルチビュープロファイルが標準化された。より自然な 3 次元映像を得るためには、多眼 3 次元画像を用いる方が良いが、符号化方式として MPEG2 マルチビュープロファイルと同じ方式を拡張した場合、莫大な符号化量となってしまう。そこで本論文では、多眼 3 次元動画の符号化方式として、周波数成分再合成法を提案する。この方式では、まず一眼分だけ MPEG2 で符号化する。そして、残りの眼はブロック平均により解像度を小さくした画像を符号化し、カットされた周波数成分を復号時に再合成する。本方式のブロック平均画像における符号化量は、MPEG2 と比較して 10% 程度である。また、3 次元ディスプレイを用いて主観評価を行ったところ、十分な画像品質を保っていることが分かった。

## A Highly Efficient Multiview Motion Picture Encoding Using Disparity Information

Hideyuki Yakiwara, Tomoyuki Hamamura, Hitoshi Aida and Tadao Saito

Graduate School of Engineering, University of Tokyo

MPEG2 multiview profile using temporal scalability for stereo image encoding has been standardized. However, using multiview picture results in more natural 3-D images. When extending MPEG2 multiview profile to include multiview picture, the bit-rate increases greatly. In this paper, we propose a Frequency component reconstruction method for multiview picture encoding. In this method, first, one view sequence is encoded using MPEG2. The remaining view sequence is compressed by block averaging and subsampling. The cut frequency component is reconstructed during decoding. By block averaging and subsampling, the resulting bit-rate is about 10% of that when using MPEG2. In addition, when using subjective evaluation, very little quality degradation is experienced.

## 1 まえがき

近年、臨場間通信会議システムへの応用等で、3次元立体テレビへの期待が高まりつつある。立体画像としては、左右2眼分の情報を持つステレオ画像を用いるものが最も簡単な方法として考えられる。ステレオ画像は、平面の画像と比較して2倍の情報量を持つので、効率的な符号化が必要であり、様々な方法が研究されている[1]～[3]。これらの方法では、動き補償に加えて視差補償を行うことで、補償の精度を上げることを基本的な技術としている。ステレオ画像の符号化方式として、MPEG2マルチビュープロファイルが標準化された[4]が、この方式も動き補償と視差補償が要素技術となっている。

ステレオ画像には、観察者が左右に移動しても対象物の見える角度が変化しないという欠点がある。より自然な3次元映像を得るためには、複数の方向から見た画像群より成る多眼3次元画像を用いる方が良い。しかし、多眼3次元画像は眼数が増えるに従い情報量が増加するので、より高能率な符号化法が求められる。筆者らは、多眼3次元画像の符号化方式として、動き補償と視差補償を用いた方法について検討を行った[5]が、各眼独立に符号化するサイマルキャスト方式と比較して80%程度にしか情報量を削減できなかった。

そこで、本論文では多眼3次元画像を高能率に符号化する周波数成分再合成法を提案する。この方式は、一眼分は普通のMPEG2で符号化する必要があるが、残りの眼については、MPEG2で符号化する場合と比較して10%程度で符号化できる非常に高能率な方式である。

## 2 実験用多眼3次元動画像の撮影

符号化シミュレーションに用いる画像として、4台のデジタルビデオカメラを並べて多眼画像(4眼)を撮影した。その画像は、以下の3種類である。

卓球 室内で卓球をしている画像。カメラ固定。

道路 道路上を車が行き交う画像。カメラ固定。

大学 大学周辺の景色。カメラ左右にパン。

これらの画像は平行撮影により撮影した。しかし、カメラの位置調節は手動で行った為、正確に平行に

はなっておらず、垂直方向にもズレがある。また、カメラ間の特性の違いにより、輝度差も生じている。実際の撮影を考えた場合、このような多少のズレは起こってしまうと思われるので、今回の実験では補正を行わずにそのまま用いた。

## 3 周波数成分再合成法

まず本節では、ステレオ画像を効率的に符号化する方式について検討を行う。

### 3.1 線形補間

立体画像において、片眼に低い鮮鋭度の画像を、もう片眼に高い鮮鋭度の画像を提示した場合、最終的に知覚される鮮鋭度は高い方のものであるということが知られている[6]。この視覚特性を利用して、片眼の解像度を落して画像を圧縮する方法が提案されている[7]。この方式は、解像度を落した画像を伝送した後、線形補間により画像を復号するものである。そこで、以下のような符号化方式でシミュレーションを行ってみる。

1. 左眼画像をMPEG2を使って符号化する。
2. 右眼画像を、8画素×8画素のブロックに分割し、そのブロックの平均値を集めることにより、解像度を落した画像を作成する。
3. 2.で作成した低解像度画像をMPEG2を使って符号化する。

以上の方法を用いることにより、右眼画像は1/64に圧縮することができる。このようにして符号化した右眼画像を、線形補間を使って画素間を補間した画像が図7である。この画像は、全体的にはけた画像となっている。線形補間による復号画像のSNRの推移を表しているのが、図3～5中の‘linear’である。また、そのSNRの平均が表1に示されている。これを見ると、この方式での復号画像のSNRが低い値であることが分かる。

### 3.2 残差成分補間

線形補間により復号した画像が、全体的にはけた画像となり、SNRも低い値になってしまうのは、プ

ロック平均を取ることにより、高周波成分がカットされてしまった為である。ここでステレオ画像においては左右画像間に強い相関があることを考えると、このカットされた高周波成分は、普通の大ききで送られてくる左眼画像にある程度含まれていると予想される。そこで、右眼画像でカットされた情報を左眼画像から再合成することを考えてみる。

右眼画像でカットされてしまった情報とは、元の画素値とその画素が含まれるブロックの平均値との差である。この差のことを、残差成分と呼ぶことにする。この残差成分を補間する方法として以下のような符号化法が考えられる。

1. 右眼画像・左眼画像において、各画素値からその画素が含まれるブロックの平均値を引いて、各画素における残差成分を求める。
2. 右眼画像の4つのブロックをまとめて16画素×16画素のマクロブロックとし、マクロブロック毎に左眼画像との残差成分におけるマッチングを行い、一番誤差が少ない部分を探す。
3. ブロック平均により圧縮した右眼画像に加えて、3.で見つけた各マクロブロックの視差ベクトルを伝送する。

ここで、4つのブロックを1つのマクロブロックとしてまとめて視差ベクトルを求めるのは、1つずつのブロック毎に視差ベクトルを伝送すると、視差ベクトルの情報だけで大きな帯域を消費してしまうからである。

この残差成分の補間による方法で符号化シミュレーションを行った。その復号画像が図8である。線形補間による復号画像と比較して、細かい部分まで見えるようになっている。しかし、画像のエッジ部分などで雑音が目につく。

図3～図5 および、表1を見ると、残差成分の補間による復号画像は、線形補間と比較して、3～4[dB]程度 SNR が向上していることが分かる。

### 3.3 DCT 係数の補間

残差成分の補間による方法では、線形補間と比較して画質が向上しているが、まだ雑音が目立つ。そこで、高周波成分の補間精度を上げる為に、ブロック

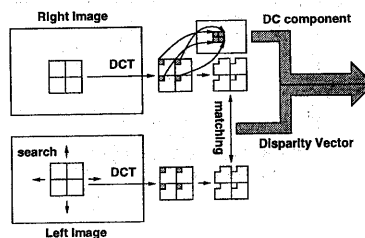


図 1: DCT 係数補間の符号化法

に DCT 演算を行った後に、その DCT 係数を補間する方法を考える。

8画素×8画素 ブロックに対して DCT 演算を行うと、ブロックの左上の方に低周波成分がきて、ブロックの右下に行くほど高い周波数成分が配置されていく。この中で、ブロックの左上隅が直流成分となる。これまでの方法で用いていたブロックの平均値とは、この直流成分にはかならない。線形補間で画像がぼけてしまったのは、直流成分を除いた63個の交流成分が削除されてしまった為である。つまり、この63個の交流成分を左眼画像から補間すれば、より綺麗な画像が復号できると思われる。そこで、右眼画像に対する符号化法として、以下のような方法を考える(図1)。

1. 右眼画像を 8画素×8画素 のブロックに分割し、それぞれに DCT 演算を行う。
2. 右眼画像の4つのブロックをまとめてマクロブロックとする。
3. 左眼画像の探索領域の中でもマクロブロックを作り、この中の4つのブロックに対してそれぞれ DCT 演算を行う。
4. 左眼画像、右眼画像それぞれで作ったマクロブロック中の4つの直流成分を除いた63×4個の交流成分同士の2乗誤差を求める。
5. 3.～4.を探索領域中で繰り返し、2乗誤差が一番小さい時を左右画像の対応点とし、その相対的位置を視差ベクトルとする。
6. 右眼画像の直流成分を集めた画像に視差ベクトルを付加して伝送する。

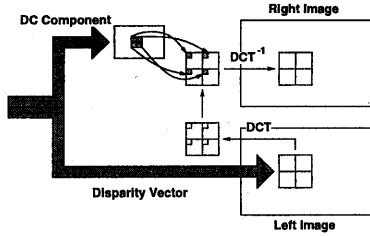


図 2: DCT 係数補間の復号法

	卓球	道路	大学
線形補間	27.42	20.05	24.91
残差成分補間	31.21	23.54	28.73
DCT 係数補間	34.52	27.37	31.63
半画素 DCT	35.42	28.65	32.93

表 1: 復号画像の SNR 平均 [dB]

このような方法で符号化したデータから右眼画像を復号するには、以下のような方法を用いる (図 2)。

- 送られてきた視差ベクトルを使って、左眼画像の該当する部分のマクロブロックを抜き出し、それに含まれる 4 つのブロックそれぞれに DCT 演算を行う。
- 送られてきた右眼画像の直流成分と、1. で DCT 演算が施されたブロックの直流成分とを入れ替える。
2. で得られた DCT 係数に対して逆 DCT 演算を行い、右眼画像を復号する。

以上の DCT 係数の補間による方法で、符号化・復号を行った右眼画像が図 9 である。この画像を見ると、残差成分の補間による方法と比較して雑音が低減していて、明らかに画質が向上している。また、図 3~図 5 および表 1 を見ると、DCT 係数の補間では残差成分の補間と比較して 3~4[dB] 程度 SNR が上がっており、予測精度が良くなっていることが分かる。

### 3.4 半画素精度の DCT 係数補間

MPEG の動き補償では、動きベクトルによるマッチングの精度を上げる為に、半画素精度での動き補

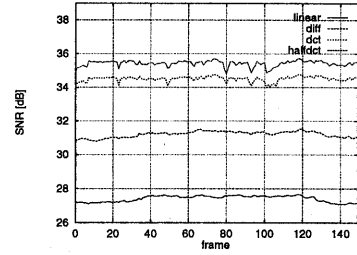


図 3: 復号画像の SNR (卓球)

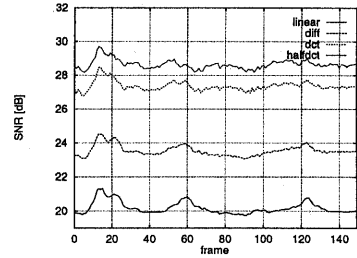


図 4: 復号画像の SNR (道路)

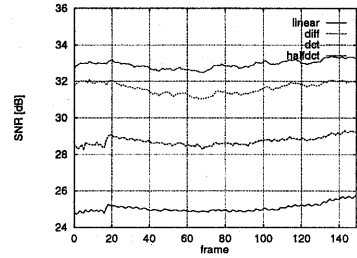


図 5: 復号画像の SNR (大学)

償が行われている。これは、まず普通に整数画素単位で動きベクトルを求めた後に、その動きベクトルを上下左右 8 近傍に半画素ずつ動かし、2 乗誤差が最小となるベクトルを探すものである。この時、予測画素が 2 画素間の時にはその 2 つの画素値の平均を、4 画素間の時にはその 4 つの画素値の平均を用いる。

この半画素精度でのマッチングを DCT 係数の補間にも応用すれば、予測精度がより向上すると思われる。具体的には、以下のような手順で符号化を行う。

- DCT 係数補間の符号化法での 4. までと同じ手順で、整数画素単位の視差ベクトルを求める。
- その視差ベクトルのまわり 8 近傍を半画素ず

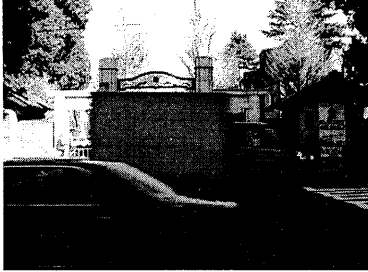


図 6: 原画像 (道路)



図 9: DCT 係数補間による復号画像



図 7: 線形補間による復号画像

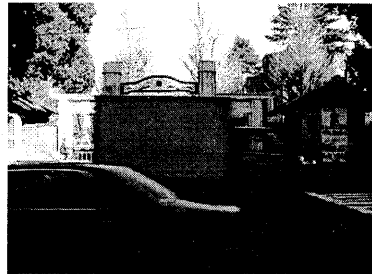


図 10: 半画素精度補間による復号画像

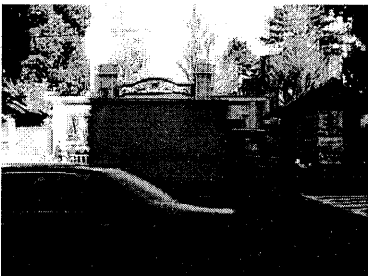


図 8: 残差成分補間による復号画像

よび表1を見ると、半画素精度の DCT 係数補間は整数画素単位のものよりも 1[dB] 程度の SNR が増加している。半画素精度の視差ベクトルを用いると、処理の量はそれほど増やさずに、少しではあるが画質を向上させることが出来る。

この半画素精度の DCT 係数の補間による画像の符号化法を、周波数成分再合成法と呼ぶことにする。以下では、この方法についてさまざまな検討を行う。

つ動かししたマクロブロックを作り、DCT 演算を行った後、周波数成分のマッチングにより 2 乗誤差を求める。

3. 1. 2. の中で一番 2 乗誤差が小さいものを半画素精度の視差ベクトルとし、伝送する。

復号に関しては、半画素精度の視差ベクトルを用いて左眼画像から予測画像を得ること以外は、DCT 係数の補間と同じである。

以上の方法で符号化・復号を行った右眼画像が図 10 である。この画像を見ると、整数画素単位の DCT 係数補間による方法と比較して、それほど目に付くような画質の向上は見られない。また、図 3～図 5 お

## 4 多眼画像への拡張

### 4.1 多眼画像への拡張における問題点

周波数成分再合成法では、1 眼分の画像だけは普通の大きさで送らなければならない。この一番綺麗な画像はできるだけ中心に位置した方が良いので、左から数えて 2 眼目を普通の MPEG2 で符号化する眼とする。すると、1 眼目と 3 眼目については、2 眼目の画像を参照画像として、これまでの右眼画像と同じ方法で符号化すれば良い (図 11)。

しかし、ここで問題となるのは 4 眼目である。4 眼目については、一番近い 3 眼目の復号画像から周

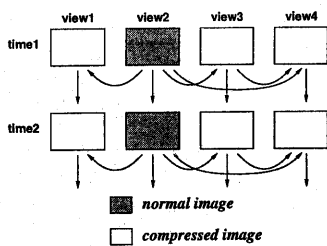


図 11: 多眼画像符号化方式

	卓球	道路	大学
2 眼目参照	33.27	27.83	32.10
3 眼目参照	35.09	27.81	31.81
2 画像からの選択	35.37	28.20	32.37

表 2: 4 眼目復号画像の平均 SNR [dB]

波数成分を再合成するか、あるいは一番綺麗な 2 眼目の復号画像から周波数成分を再合成するかの 2 通りが考えられる。そこで、2 眼目を参照画像とした場合と 3 眼目を参照画像とした場合についてシミュレーションを行った。その結果の平均 SNR が表 2 に示されている。これを見ると、卓球では 3 眼目を参照した方が良くなっているが、逆に大学では 2 眼目を参照した方が良くなっている。道路では、どちらを参照した場合もほぼ同じ SNR となっている。

以上のように画像によって結果にばらつきがあるのは、カメラから被写体までの距離が起因になっていると思われる。例えば、被写体が近い距離にある卓球では、カメラの位置により映っている画像にある程度のズレが生じる為、一番近い 3 眼目を参照画像にした方が有効である。一方、被写体までの距離が遠い大学では、カメラの位置が少々変わっても映っている映像はそれほど変わらないので、一番綺麗な画像である 2 眼目からの参照が有効である。

#### 4.2 2 つの画像の選択による符号化法

4 眼目については、2 眼目と 3 眼目のどちらを参照画像とした方が良いかは、画像によって異なってしまうことが分かった。そこで、マクロブロック毎に 2 眼目と 3 眼目それぞれを対象画像として予測を行い、予測精度の良い方を参照画像として用いることにする。

	卓球	道路	大学
1 眼目	342234	302381	353327
3 眼目	337344	301146	350343
4 眼目	360119	317705	381894

表 3: 可変長符号化による視差ベクトルの符号化量

このような方法により、4 眼目を符号化・復号した画像の SNR の平均が表 2 に示されている。これを見ると、当然ながら 2 眼目からの選択による方法が一番良くなっている。しかし、2 眼目からの選択を行う場合には、マクロブロック毎にどちらの参照画像を選択したのかを表すビットを付加しなければならず、少々符号化量が増えてしまう。

## 5 視差ベクトルの符号化量

周波数成分再合成法では、解像度を落した直流成分画像に加えて視差ベクトルを伝送しなければならない。今回の実験では、視差ベクトルの探索範囲を  $-15 \sim 15$  としており、さらに半画素精度のマッチングも用いている為に、マクロブロック一つにつき無圧縮で 12[bit] が必要となる。画像一枚の大きさは  $640 \times 480$  なので、一秒間では

$$\frac{640}{16} \times \frac{480}{16} \times 12 \times 30 = 432000[\text{bit/s}] \quad (1)$$

の符号化量となる。

これは、全体の符号化量の中でも大きなウェイトを占めるので、できるだけ圧縮する必要がある。そこで、視差ベクトルについては、MPEG2 の動きベクトルの符号化で用いられているような可変長符号化を用いることにする。これは、当該マクロブロックと一つ前 (左隣) のマクロブロックの視差ベクトルの差分を取り、それを可変長符号 (Variable Length Code) により効率的に符号化する方式である。

可変長符号化により視差ベクトルを符号化した結果が、表 3 である。これを見ると、全ての場合において無圧縮時よりも符号化量が削減できている。

## 6 主観評価

### 6.1 修正 EBU 法

文献 [8] では、修正 EBU 法と呼ばれる画像主観評価法が提案されている。この方式では、図 12 に示

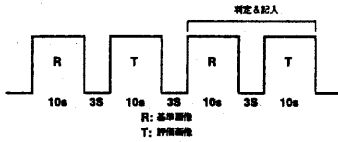


図 12: 修正 EBU 法における動画の提示方法

評点	評価用語
5	(劣化が) わからない
4	(劣化が) わかるが気にならない
3	(劣化が) 気になるが邪魔にならない
2	(劣化が) 邪魔になる
1	(劣化が) 非常に邪魔になる

表 4: 修正 EBU 法で用いられる評価用語

すように、基準となる画像と評価画像とを 2 回繰り返して提示し、その評価画像が基準画像に対してどのように被験者に知覚されるかを、5 段階劣化尺度を用いて評価する。5 段階劣化尺度で用いられる評価用語は、表 4 に示すものである。今回の主観評価は、この修正 EBU 法を用いて行った。

## 6.2 主観評価実験の条件

今回行った主観評価実験の条件は、表 5 に示すものである。モニターには、バララックスバリア方式の 3 次元ディスプレイを用いた。このディスプレイの適視距離が 50 ~ 60cm である為、視距離もこの距離とした。また、実験に用いた多眼画像がそれぞれ 5 秒の動画である為、画像の提示時間を 5 秒に短縮して実験を行った。

今回の実験では、周波数成分再合成法の性能を評価する為に、線形補間による符号化画像に対しても評価を行った。線形補間では、必ず片方には綺麗な画像を提示する必要があるため、1 眼目と 3 眼目を圧縮後に線形補間を行う画像、2 眼目と 4 眼目を MPEG2 でそのまま符号化する画像とした。また、修正 EBU 法での基準画像には、4 眼それぞれを独立に符号化するサイマルキャスト方式の画像を用いた。各方式で MPEG2 で符号化する画像は 6Mbps、ブロック平均により低解像度にした画像は 300kbps のビットレートで符号化した。この場合、周波数成分再合成法により符号化した画像では、視差ベクトルは平均で 350kbps なので、合計で約 650kbps となる。

評価法	修正 EBU 法
視距離	50 ~ 60cm
評定者	30 名
テスト画像	5 秒間の 3 次元動画 3 種類
モニター	10.4 インチ 3 次元ディスプレイ (三洋電気製 THD-10PN3)

表 5: 主観評価実験の条件

	卓球	道路	大学
サイマルキャスト	24.01	24.03	24.01
線形補間	12.30	12.32	12.31
周波数成分再合成	7.99	7.87	8.00

表 6: 評価画像の符号化ビットレート [Mbps]

4 眼分の多眼画像を用いた場合、3 方向から立体像を観察することができる。そこで、今回の評価実験では、これらの画像について以下の 3 方向から見た場合に対して行った。

**view1-2** 左眼で 1 眼目の画像を、右眼で 2 眼目の画像を見る場所。

**view2-3** 左眼で 2 眼目の画像を、右眼で 3 眼目の画像を見る場所。

**view3-4** 左眼で 3 眼目の画像を、右眼で 4 眼目の画像を見る場所。

## 6.3 主観評価結果

以上のような方法により主観評価実験を行った結果が、図 13~図 15 である。これらを見ると、全ての画像、全ての方向において線形補間方式よりも周波数成分再合成方式の方が画質が優れていることが分かる。

また、周波数成分再合成方式をサイマルキャスト方式と比較した場合、view1-2 や view2-3 では、「劣化が分かるが気にならない」という 4 点の評点よりも平均が上回っている。しかし、view3-4 では、道路で 4 点を僅かに下回り、卓球と大学では「劣化が分かるが気にならない」と「劣化気になるが邪魔にならない」の間である 3.5 点程度となっている。これは、view3-4 では左右両眼とも、周波数成分再合成により作成した画像であり、両眼ともに劣化した画像を見るために、立体映像としても劣化がある程度気になってしまうと考えられる。

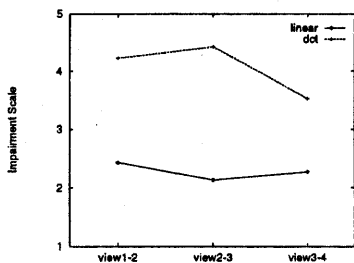


図 13: 主観評価結果 (卓球)

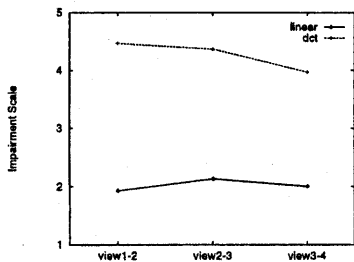


図 14: 主観評価結果 (道路)

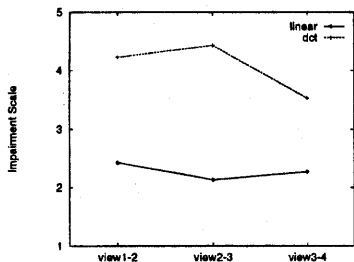


図 15: 主観評価結果 (大学)

周波数成分再合成法で低い評価が付けられる場合には、画像のちらつきが気になったという意見が大部分を占める。この画像のちらつきは、直流成分画像をそのまま用いた場合には、全く見られない。したがって、直流成分を MPEG2 で符号化することによる劣化が、復号画像がちらつく原因だと言える。

## 7 むすび

本論文では、多眼 3 次元動画画像の高効率な符号化法として、周波数成分再合成法を提案した。この方式は、一眼分だけ綺麗な画像を伝送すれば、残りの眼については MPEG2 で符号化する場合と比べて 10% 程度で符号化できる為、多眼画像の眼数を増加させ

ても、それほど消費帯域は大きくならない。また、主観評価を行った結果、良好な品質が得られていることが分かった。しかし、両眼ともに周波数成分再合成により復号した画像の場合には、知覚品質が少し落ちてしまう。この主な原因は、直流成分画像の MPEG2 符号化による劣化である。したがって、今後はこの直流成分を効率的に可逆符号化する方法、あるいは非可逆符号化でも、復号画像でちらつきをあまり感じさせない方法を検討していく予定である。

## 参考文献

- [1] W.A. シュップ, 安田, “視差補償および動き補償を用いたステレオ動画画像のデータ圧縮”, PCSJ88, pp.63-64, 1988
- [2] 泉岡 生見, 渡辺 裕, “視差補償予測を用いたステレオ動画画像の符号化”, 信学技報 IE89, pp.1-7, 1989
- [3] Belle L. Tseng, Dimitris Anastassiou, “Compatible Video Coding of Stereoscopic Sequences using MPEG-2’s Scalability and Interlaced Structure”, HDTV’94, pp.6-B-5, 1994
- [4] ISO 13818-2 AMD3 Multiview Profile, 1996
- [5] 八木原 英之, 相田 仁, 齊藤 忠夫, “MPEG2 の拡張による多眼 3 次元動画画像符号化方式の検討”, 情処学 55 回全大 3-521, 1997
- [6] T.Mituhashi, “Subjective image position in stereoscopic TV systems - Considerations on comfortable stereoscopic images -”, In Proc. SPIE, No.2179, pp.54-60, 1994
- [7] Micheal G.Perkins, “Data Compression of Stereopairs”, IEEE Trans. Communications, Vol.40, No.4, pp.684-696, 1992
- [8] 成田 長人, “画像品質の主観評価法に関する一考察”, NHK 技研 R&D, No.34, 1994