

注視点情報を用いた映像メディア符号化アーキテクチャの検討

笠井 裕之 児玉 明 前田 鏡二 富永 英義

早稲田大学 理工学部 電子通信学科

〒169 東京都新宿区大久保 3-4-1

kasai@tom.comm.waseda.ac.jp

本稿では、高能率映像メディア符号化アーキテクチャとして視覚情報を用いた符号化アーキテクチャを提案し、その概念について述べる。特に今回視覚情報として、人間の視力特性に着目し、提案アーキテクチャにより提供される画像の視覚効果について検証する。その結果、注視点に情報を多く割り当てた方が視覚的に高画質化が図れることを確認した。次に、本提案符号化アーキテクチャの例として、2つの手法を用いた。まず既存の符号化制御方式への適用手法には、視覚特性関数を用いて検討した。また、階層符号化方式として視覚特性に応じたスケーラビリティを提案し、利用者の視覚情報に基づいた要求機能を提供する符号化方式を検討する。シミュレーション実験の結果、それぞれの手法において提案アーキテクチャを実現できた。特にスケーラビリティでは、人間の視覚情報を用いた映像サービスのアプリケーションならびにその特徴を示し、その有効性を示した。

A Study on Video Coding Architecture using Gaze Point Information

Hiroyuki KASAI Mei KODAMA Keiji MAEDA Hideyoshi TOMINAGA

Dept. of Electronics and Communication Engineering, WASEDA University

3-4-1 Ohkubo Shinjuku-ku, Tokyo, 169 JAPAN

kasai@tom.comm.waseda.ac.jp

In this paper, we propose video coding architecture using gaze point information as high efficiency architectures and explain the concept. Especially, we focus on eyesight characteristic. We verify visual effect given by proposal architecture. In consequence, the more coding bits are allocated to gaze point, the better visual image quality is. This architecture is realized by using coding control method and scalability scheme. This method is used by eyesight characteristics functions. On the other hand, scalability one offers visual requirement functions. Finally, we show visual service application and the feature in this scalability.

1. はじめに

画像符号化を行なう際、定められた画像サイズに対し、定められた符号量の中で可能な限り視覚的に劣化のない符号化画像を作成することが要求される。一つの手法として、局所領域ごとの符号化制御方式が検討され、動画像標準符号化方式(以下、MPEG2)にも採用されている⁽¹⁾。しかし現在の手法では画像フレーム全体に対する視覚的な重みづけの処理しかしていないため、人間の視覚特性を十分考慮したものであるとは言い難い。そこで本研究では、さらなる人間の視覚的な効果を利用した高能率符号化アーキテクチャとして、視線の位置(以下、注視点と呼ぶ(Gaze Point))における高解像度画像を提供する符号化方式について検討する。

本稿では、まず人間の視野内での視力の分布特性を述べる。次に人間の視覚情報を用いた映像メディア符号化アーキテクチャの概念について述べ、その視覚効果について検証する。さらに提案方式の実現手法として、既存の符号化制御方式へ適用する手法と階層符号化方式の2種類の手法について検討する。人間の視覚特性として視力特性に着目する。今回は検証方法としてMPEG2を用いて行なうことにする。

2. 人間の視覚情報を用いた映像メディア符号化アーキテクチャの提案

2.1 視野内における人間の視力特性

人間の眼を動かさずに一度に見える範囲を視野というが、その範囲は上下左右で60度の広がりである。しかし、視野内で一律の視力があるわけではない。網膜上の視力の分布を図1⁽²⁾より引用)に示す。視力は網膜上の1度20分の範囲で最もよく、ここから離れるにしたがい急激に減少する。すなわち、人間は視野内の注視点部分を詳細に視ている(中心視)であり、周辺視は低解像度で済ませている。つまり人の眼が一望で受容できる画素数はそれほど多くはないと言える。

2.2 視力特性を利用した符号化方式の概念

2.1のような人間の視覚特性から、人間が画像を見る際、画像全体の全ての情報を得ることは不可能である。周辺部は見える部分であっても、認識することはできない。人間が得ることのできる情報は、注視点とその近傍付近に集中しており、周辺部分の情報はあまり取得出来ず、その部分の情報は破棄されている。よって周辺部には注視点ほどの多くの情報を与える必要はないと考えられる。

そこで、人間の画像上の注視点部分に近いほど高

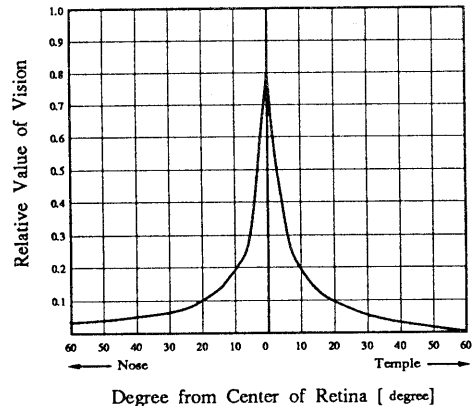


図1: 網膜の位置と視力の関係

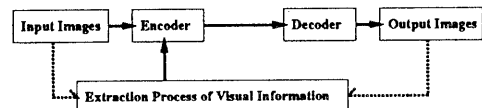


図2: 提案符号化アーキテクチャの概念

解像度化された画像を提供する映像メディア符号化アーキテクチャを提案する。符号化の際、人間の視覚情報を抽出し、その情報を外部入力とする。本方式では、人間の画像上の注視点部分により多くの情報を割り当て、周辺部の情報を削減することによって高能率化を図る。提案方式の概念図を図2に示す。

2.3 視力特性を用いた符号化方式の検証

視覚特性に応じた符号化再生画像に対する視覚効果の検証のため、シミュレーション実験を行なう。

画像上の注視点の位置に最も優先度があり、注視点から離れる程、優先度が低くなっていく。そこで本方式では図3のように、優先度に対して優先度の高いものからクラスA、クラスB、・・・というようにクラス分けをする。以下、クラスの範囲(半径 R_K)を求める。

図1において、網膜の中心からの視角の度数を x としたときの視力の相対値を示す関数を視力特性関数 $f(x)$ とする。ただし図1において $x=0$ (Y軸)について線対称であると近似して、 $x \geq 0$ の値を用いて、以下述べる。

□ 注視点情報を用いた映像メディア符号化アーキテクチャの検討

また、以下の図3のようにパラメータを設定するとクラスの半径 R_K は、式(1)により求められる。

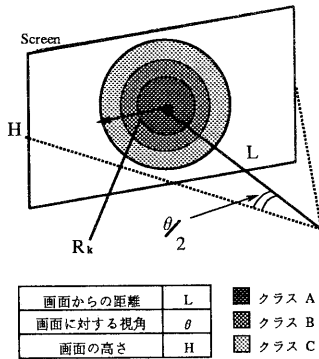


図 3: クラス分けとパラメータ設定条件

$$R_K = L \times \tan\left[f^{-1}\left\{1 - \frac{1 - f(\theta/2)}{2(n-1)}(2k-1)\right\}\right] \quad (1)$$

ここで画像システムの性能や品質を評価する場合、目的や状況に最も適した評価方法、観視条件を選択する必要がある。そこで今回は、画像評価法として、CCIR 勧告 500⁽³⁾ に基づき最適視距離として 4H(H:画面の高さ)を用いて画像評価を行なう。ただし、今回画像を評価するディスプレイは、アスペクト比 4:3 のものを使用するので画面に対する視角 $\theta=20$ 度(計算省略)を用いる。

2.3.1 MPEG2 の利用

実際の視力特性に応じた処理画像の作成方法として MPEG2 を用いる。MPEG2 での符号化単位はマクロブロック (16[pel]x16[line], 以下 MB と呼ぶ) であるので、本方式では図4のように、最も優先度の高いクラス A に正方形の MB 群を割り当て、その MB 群を中心として同心円上に正方形のドーナツ型の MB 群をクラス B,C,D に割り当てるといったクラス分けを考える。今回は簡略化のため、一例として 4 つのクラス (n=4) の場合について考える。勿論、n 個のクラスにも拡張可能である。

そこで式(1)から、クラスごとの正方形 MB 群の一边のマクロブロック数 mb_K を算出する。今回用いる入力画像は 704x480 画素の画像であるので、 mb_K は式(2)のように求められる。

$$mb_K = 66 \times \frac{R_K}{H} \quad (2)$$

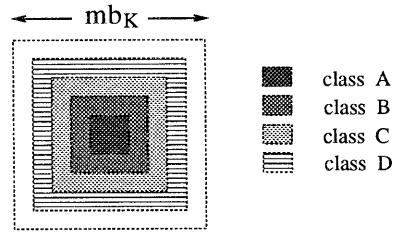


図 4: 本方式におけるクラス分け

これより、クラス毎の一边のマクロブロック数 mb_A, mb_B, mb_C, mb_D は、それぞれ 6, 8, 22, 44 となる。ただし、2.1 で述べたような網膜上の特性から、最も優先度の高いクラス A の半径は視角 1 度 20 分の値を用いて算出した。

今回のシミュレーション実験では、視力特性に応じた処理画像と非処理画像との比較検証を、MPEG2 における量子化ステップ制御パラメータを変動させることにより行なう。

MPEG2 では視覚的重要度に応じて DCT 係数の量子化ステップを領域毎に変動させることにより画質を変えることができる。このことを利用して、本方式では、視覚的に重要度の高い部分、すなわち注視点部分には量子化ステップを小さくし、それに対して、注視点から離れている部分を重要度の低い部分として量子化ステップを大きくする手法を用いて、提案方式の視覚効果検証を行なう。MPEG2 における量子化式は式(3)に示される。

$$QAC(i,j) = \frac{16 \times ac(i,j)}{2 \times MQ \times w_N(i,j)} \quad (3)$$

ここで、(i,j) はブロック中の水平、垂直方向の位置、 $ac(i,j)$ は 8x8 ブロック中の DCT 係数、 $QAC(i,j)$ は量子化係数である。

そこで今回のシミュレーションでは、量子化ステップ制御パラメータとして、式(3)の分母の MQ を使用し、各クラスごとに MQ の値を変化させる方式を用いる。

2.3.2 MQ の満たすべき条件

ここでは、検証実験においてクラスごとに変動させる MQ の満たすべき条件を、視力特性の図 1 から導出する。

視力特性に応じた処理画像と非処理画像のフレーム全体の符号量をほぼ同じにするという条件から、

クラス A の値を基準として次の式が満たされたされなければならない。

$$\sum_{K=A} \frac{\overline{f_A}}{\overline{f_K}} \times S_K = \frac{\overline{f_A}}{\overline{f(x)}} \times \sum_{K=A} S_K \quad (4)$$

ただし S_K はクラス K の面積、 $\overline{f(x)}$ はフレーム全体の視力の相対平均値、 $\overline{f_K}$ はクラス K における視力の相対平均値である。

ここで、今回 S_K に対応する係数 $\overline{f_A}/\overline{f_K}$ を便宜上 MQ の線形制御に用いる。各クラスの MQ の比率を算出すると、 $MQ_A : MQ_B : MQ_C : MQ_D$ は 1.0:1.7:2.6:4.1 となる。ただしクラス A において視力が最も良いことから $\overline{f_A} = 1$ として導出した。

2.3.3 シミュレーション実験

mb_K と MQ_K を用いて視力特性に応じた処理画像と非処理画像を作成する。そのときのシミュレーション実験条件を表 1 に、シミュレーション結果を表 2、表 3 に示す。また、視力特性に応じた処理画像と非画像の例を図 5 に示す。

表 1: 符号化実験条件

シーケンス	Mobile & Calendar
画像	CCIR Rec.601 4:2:0
フォーマット	輝度信号:704[pe]x480[line] 色差信号:352[pe]x240[line]
比較方式	MQ 変動 MQ 固定
評価項目	同符号量でのフレーム全体の SN 比 同符号量でのクラス別の SN 比 作成画像に対する主観評価

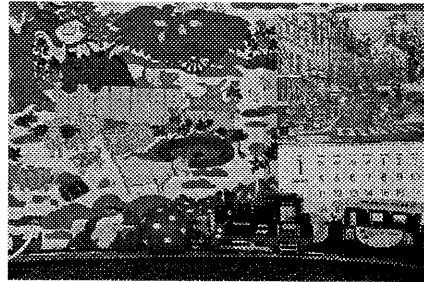
表 2: 処理画像の SN 比 (符号量 412685[bits])

	SNR			MQ
	Y[dB]	Cb[dB]	Cr[dB]	
全体	19.94	27.90	27.17	
クラス A	27.83	30.11	31.83	5*1.0
クラス B	22.93	28.68	26.11	5*1.7
クラス C	20.66	28.82	27.14	5*2.6
クラス D	19.41	27.41	27.11	5*4.1

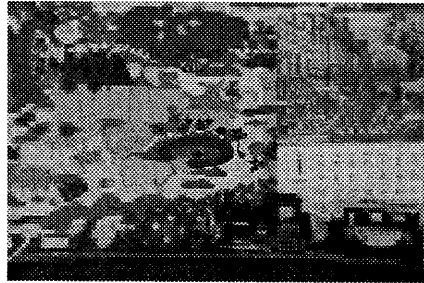
視力特性に応じた処理画像と非処理画像とでは、フレーム全体の SN 比は非処理画像の方が良い値であるが、クラス毎の SN 比の変化は視力特性に応じて処理を行った効果が顕著に現れている。また、主

表 3: 非処理画像の SN 比 (符号量 414716[bits])

	SNR			MQ
	Y[dB]	Cb[dB]	Cr[dB]	
全体	19.97	27.89	26.88	
クラス A	19.80	24.16	21.99	5*3.2
クラス B	19.50	26.04	23.92	5*3.2
クラス C	19.77	28.13	26.32	5*3.2
クラス D	20.11	28.11	27.80	5*3.2



(a) 原画像



(b) 処理画像

図 5: 視力特性に応じた処理画像例

観評価では処理画像は非処理画像に比べ、画像の端ではブロック歪みが、クラスとクラスの境界部で境界歪みが観察された。しかし、注視部分を視た場合、周辺部でのそのような歪みはまったく問題がないことが観察された。

同じビット量にも関わらず、注視点に情報を多く割り当てた分、視覚的に高画質化が図られた。このことは、正に人間の視覚特性を利用したものに他ならない。つまり、視力特性を利用した提案符号化アーキテクチャの視覚効果が示された。

□ 注視点情報を用いた映像メディア符号化アーキテクチャの検討

3. 提案方式の実現手法

2.により、視力特性を利用した映像メディア符号化アーキテクチャの有意性が確認された。そこで、本提案方式の実現手法として、既存の符号化制御方式に適用する手法と階層符号化方式を用いた手法を検討する。

3.1 符号化制御方式を用いた実現手法

視力特性を用いた符号化アーキテクチャの実現手法として、符号化制御、その中でも適応的な量子化制御を利用する手法を検討する。この手法は、決められた符号レートの中で符号化効率を向上させながらも、その中で視覚的な重みをかけた符号化を実現することを目的とする。

MPEG2では、画像情報の中でも輝度情報を用いた符号化制御方式が利用されている。本方式では、図1に示される視力関数 $f(x)$ を人間の視力特性のパラメータとして利用する。本研究ではMPEG2方式としてTest Model(以下, TM⁽⁴⁾)を利用する。まずTMにおける量子化制御法を簡単に説明する。

TMにおける符号化制御はMB毎に適応的にパラメータを変えることで行う。符号化時のMQの決定には、輝度値Yの分散値から得られる act_j の変数を利用している。そこで、本方式では、分散値を用いるかわりに、視力特性関数 $f(x)$ を式(5)のように用いる。

$$act_j = \frac{1}{f(x)} \quad (5)$$

この方式は、 $f(x)$ が1に近いとき、すなわち注視点に近いときは、そのMBが重要であると考えられるので、量子化歪みが目立つとして量子化幅を小さくする。逆に $f(x)$ が1に比べて小さいとき、すなわち注視点から離れているときは、そのMBがあまり重要でないと考え、量子化幅を大きくする。

3.1.1 シミュレーション実験

視力特性関数を符号化制御に用いた方式の効果を検証するため、TM方式との比較シミュレーション実験を行なった。シミュレーション実験条件を表4に示す。また、Ballet, mobilの実験結果をそれぞれ、表5,6と表7,8に示す。

TM方式と視力特性関数を用いた方式の比較では、それぞれの視覚特性を量子化制御に適用した効果が現れている。視覚特性関数の利用により汎用的符号化アーキテクチャに沿った符号化方式が容易に実現できる。つまり、視力特性を量子化制御に適用することによって、今回の視力特性を利用した符号化アーキテクチャが実現できることが分かる。

表 4: シミュレーション実験条件

方式	視力特性関数方式, TM方式
シーケンス	Mobile & Calendar, Ballet
符号化画像数	30 フレーム
予測・符号化	フレーム予測, フレームDCT
画像フォーマット	CCIR Rec. 601 4:2:0 輝度信号:704[pe]x480[line] 色差信号:352[pe]x240[line]
符号化レート	9[Mbps]
評価項目	原画とのS/N比 符号化ビット量

表 5: 視力特性関数方式のSN比 (Ballet)

クラス	SNR		
	Y[dB]	Cb[dB]	Cr[dB]
A	38.46	44.80	42.68
B	38.76	44.76	43.06
C	38.81	45.76	44.97
D	41.12	46.64	46.48

表 6: TM方式のSN比 (Ballet)

クラス	SNR		
	Y[dB]	Cb[dB]	Cr[dB]
A	39.48	46.36	44.86
B	39.76	46.22	44.99
C	39.84	46.61	46.05
D	41.80	47.09	46.98

表 7: 視力特性関数方式のSN比 (mobil)

クラス	SNR		
	Y[dB]	Cb[dB]	Cr[dB]
A	33.77	35.93	36.95
B	33.06	35.92	37.00
C	31.79	35.90	36.09
D	30.44	35.49	35.64

表 8: TM方式のSN比 (mobil)

クラス	SNR		
	Y[dB]	Cb[dB]	Cr[dB]
A	31.40	34.21	35.71
B	31.68	34.95	36.04
C	30.18	35.32	35.31
D	29.84	35.42	35.52

3.2 階層符号化方式を用いた実現手法

視力特性を用いた符号化方式の実現手法として、階層符号化方式について検討する。この階層符号

化方式では、注視点における高解像度化のため、スケーラビリティ機能を利用する。スケーラビリティ機能は、多種多様な情報の中から利用者が必要とする画像情報だけを取得することを可能にする機能である。ここでは、注視点における視力特性に応じた高解像度化を可能にするスケーラビリティを提案し、その有効性を検討する。その概念図を、図6に示す。

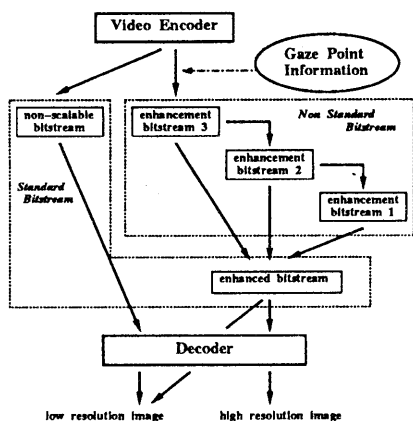


図 6: 局所的高解像度スケーラビリティの概念図

このスケーラビリティは、任意の注視点およびその近傍部分において多段の階層化を作成することにより、人間の視覚情報と連動した局所的な高解像度化を実現することができる。ここではベースとなる画像情報に、局所的な高解像度情報を付加することによって、視力特性を利用した本提案アーキテクチャ方式を実現する。図7に階層化例を示す。ただし、階層数を4とした。

3.3 シミュレーション実験

提案スケーラビリティを用いて、提案アーキテクチャの実現性の検証のためシミュレーション実験を行なう。本方式では、SNR スケーラビリティを基準にし、図8に示す符号化器を用いた。

シミュレーション実験条件を表9に、実験結果を表10,11に示す。

階層が上につれ各クラス毎のSN比は向上し、4つの階層が実現された際には、SN比は視力特性に基づいた値となっている。この結果は、図7の階層化例の通りである。つまり、ここで提案したスケーラビリティを用いることにより、視力特性を利

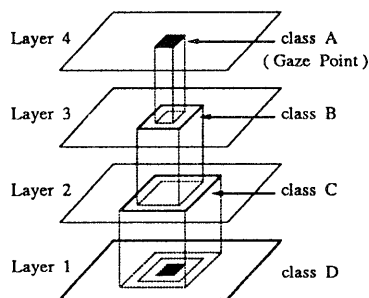


図 7: 局所的高解像度スケーラビリティ階層化例

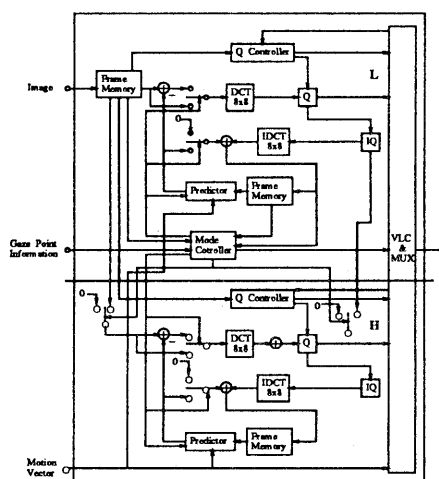


図 8: 局所的高解像度スケーラビリティの符号化器

表 9: シミュレーション実験条件

シーケンス	Mobile & Calendar, Ballet
符号化画像数	30 フレーム
予測・符号化	フレーム予測, フレーム DCT
画像フォーマット	CCIR Rec.601 4:2:0 輝度信号:704[pe]x480[line] 色差信号:352[pe]x240[line]
符号化レート	階層 1:2.3[Mbps], 階層 2:1.2[Mbps] 階層 3:0.3[Mbps], 階層 4:0.2[Mbps]
評価項目	各階層画像と原画との S/N 比 符号化ビット量, 主観評価

用した映像メディア符号化アーキテクチャが実現できていることが分かる。

□ 注視点情報を用いた映像メディア符号化アーキテクチャの検討

表 10: 階層毎の各クラスの SN 比 (Ballet)

階層	クラス	SNR		
		Y[dB]	Cb[dB]	Cr[dB]
1	A	32.13	40.64	37.29
	B	32.24	40.49	37.28
	C	34.08	41.47	39.51
	D	35.63	41.36	40.56
2	A	36.82	42.59	39.83
	B	36.98	42.66	39.97
	C	38.50	43.38	42.08
	D	37.42	41.94	41.56
3	A	39.12	44.04	41.67
	B	39.27	44.07	41.89
	C	38.84	43.58	42.42
	D	37.52	41.99	41.67
4	A	40.97	45.32	43.38
	B	40.23	44.75	42.82
	C	38.94	43.66	42.55
	D	37.55	42.01	41.71

表 11: 階層毎の各クラスの SN 比 (mobil)

階層	クラス	SNR		
		Y[dB]	Cb[dB]	Cr[dB]
1	A	24.89	28.82	29.84
	B	24.91	29.11	29.90
	C	23.99	30.55	30.20
	D	23.78	30.84	30.67
2	A	28.14	31.30	31.98
	B	28.16	31.67	32.25
	C	27.48	32.95	32.95
	D	24.70	31.63	31.59
3	A	30.87	34.00	35.14
	B	30.88	34.43	35.27
	C	27.71	33.32	33.28
	D	24.74	31.72	31.67
4	A	31.23	34.97	36.05
	B	30.98	34.94	35.65
	C	27.70	33.35	33.30
	D	24.74	31.73	31.68

階層符号化方式による手法には、次のような特徴が挙げられる。

1. 狭帯域ネットワーク上で画像伝送を行なう場合、局所的な高解像度画像だけを伝送することにより、高速かつ効率的に利用者の要求を満たすことが可能となる。

2. 画像上の全ての点における高解像度情報を階層化してデータベース化しておくことにより、多数ポイントでの高解像度画像を作成できる。

また次のアプリケーションサービスが考えらる。

1. 注視点抽出が可能である⁽⁵⁾ことを利用して、画像を眺める人の目の視線移動に連動してリアルタイム的に中心部のみを高解像度に、周辺部は低解像度にすることが可能になる。

2. 人間が取得する視覚情報をデータベース化する際のアーキテクチャとして利用できる。

4. まとめ

本稿では、人間の視力特性を利用して、注視点における高解像度化を実現することのできる映像メディア符号化アーキテクチャを提案し、その有意性を検証した。さらに、符号化制御方式および階層符号化方式による実現手法を検討した。この結果、符号化制御方式では一定符号化レートのもとで視力特性を利用した制御により、提案符号化アーキテクチャを実現することができた。また階層符号化方式でも視力特性に応じた高解像度情報を階層化することにより実現することができた。

今後の課題として、さらなる階層アーキテクチャの検討が必要である。また今回は視覚情報の内、視力特性を用いて検討を行なったが、視覚情報には視力特性だけでなく他にも様々な特性が存在する。そのような特性を統合的に扱うアーキテクチャの検討が必要である。

参考文献

- [1] ISO-IEC 13818-2, Draft International Standard: "INFORMATION TECHNOLOGY - GENERIC CODING OF MOVING PICTURES AND ASSOCIATED AUDIO Recommendation H.262" (1994).
- [2] 原島 博 監修: "画像情報圧縮", オーム社 (1993).
- [3] 成田 長人: "CCIR 勧告 500-5 画質評価法", TV学誌, 47, 7, pp. 979-982 (1993).
- [4] ISO-IEC JTC1, MPEG 93/457: "Test Model 5" (1993).
- [5] Applied Science Laboratories: "EYE TRACKING SYSTEMS HANDBOOK" (1994).