

## テクスチャの再現性を考慮した H.264 符号化モード選択方式

吉野 知伸 内藤 整 酒澤 茂之 小池 淳

株式会社 KDDI 研究所 〒356-8502 埼玉県ふじみ野市大原 2-1-15

E-mail: {to-yoshino, sei, sakazawa, koike}@kddilabs.jp

あらまし 従来の IPTV サービスでは、低ビットレートでの運用が求められ、HDTV に対して許容される帯域は高々 10Mbps 程度である。従って HDTV コンテンツの配信用途においては、高い符号化効率を維持する目的から H.264/AVC High Profile の採用が有力である。しかしながら、低ビットレートで安定した伝送品質を確保するためには、圧縮効率の更なる改善が必要となる。H.264 符号化における符号化効率の改善要素として、マクロブロック単位の符号化モード選択が挙げられる。従来、符号化モード選択手法として R-D 最適化法が用いられる。しかしながら、同手法により選択された符号化モードは必ずしも主観画質を向上させるものではなく、特にテクスチャの再現性に関して不適切な場合がある。したがって、主観画質の低下を招く符号化モードが選択されることを抑止する手法が強く求められる。以上の背景から、本研究では HDTV に対する 10Mbps 以下の低ビットレート符号化を対象として、主観画質の改善を目的とした符号化モード選択手法の提案を行う。提案手法では、局所領域における画素値の類似性およびエッジ成分の再現性を検知する新たな評価尺度を備え、これに基づいたモード選択を行うことでテクスチャ再現性を追及するものである。主観評価実験により従来手法に対して主観画質が改善することを確認し、MOS(Mean Opinion Score)が最大で 1.1 改善することを確認した。

キーワード H.264/MPEG-4 AVC, 符号化モード選択, HDTV, 低レート符号化, 主観画質

## Advanced H.264 Coding Mode Decision Scheme Effective for Texture Reality Improvement

Tomonobu YOSHINO Sei NAITO Shigeyuki SAKAZAWA and Atsushi KOIKE

KDDI R&D Laboratories Inc. 2-1-15 Ohara, Fujimino-shi, Saitama, 356-8502 Japan

E-mail: {to-yoshino, sei, sakazawa, koike}@kddilabs.jp

**Abstract** In typical IPTV network, an available bandwidth for broadcasting contents is less than 10Mbps. Therefore, H.264/AVC High Profile is generally adopted accounting for an advantage in coding efficiency. However, further improvement of coding efficiency is required to maintain the stable broadcasting quality under low bit-rate transmission. The macroblock coding mode decision is regarded as one of the key features to improve the coding efficiency of the H.264 encoding. In the conventional methods, macroblock coding mode is usually decided based on R-D optimization criteria. However, the selected mode is not necessarily appropriate to enhance the subjective picture quality especially for texture reality. Therefore, an advanced coding mode decision scheme that prevents to select the coding mode that causes the subjective picture quality degradation is strongly required. From this perspective, we propose an advanced mode decision scheme that improves the subjective picture quality for HDTV coding even under a low bit-rate less than 10Mbps. The proposed scheme selects a coding mode based on a new evaluating measure that detects the similarity of pixel values and the reality of edge component in order to pursue the texture reality. Experimental results also confirmed that our proposed scheme contributes to the subjective picture quality improvement, and the MOS(Mean Opinion Score) gain reaches 1.1 at the maximum case.

**Keyword** H.264/MPEG-4 AVC, Coding mode decision, HDTV, low bit-rate coding, subjective picture quality

## 1. はじめに

IPTV サービスでは、伝送レートの低ビットレート化に関する要求が強く、10Mbpsを下回るレートにおいてHDTVを安定的な品質で配信することが求められる。したがって、HDTVを配信するためには高い符号化効率を実現する符号化方式が求められ、H.264 High Profileの採用が有力である。しかしながら、低ビットレート環境において安定的にサービス品質を維持するには更なる画質の改善が必要である。

H.264 符号化で高い符号化効率を得られる要因の一つに、マクロブロック(MB)単位での符号化モード適応選択が挙げられる。H.264 符号化では、入力信号の特徴に応じた符号化を可能とする様々な符号化モードが定義されており、同符号化モードの適応選択により高精度の予測値を得ることができる。すなわち、H.264 エンコーダの性能は、符号化モード選択手法の成否に大きく左右される。H.264 リファレンスエンコーダ(JM)[1]では、MB 単位の符号化モード選択手法として、R-D 最適化法が採用されている。同手法は、決められた符号量の下で客観画質(PSNR)を最大化する符号化モードを選択する。しかしながら、同手法は単一のMB のみにおける客観画質を評価するため、同手法の選択結果は必ずしも主観画質を最大化するものではなく、さらに近傍MBも含めた局所領域における主観画質の低下を検知することができない。これに起因して主観画質の低下につながる典型例として、テキストチャ再現性の低下があり、特に低ビットレート符号化において顕著である。

以上の背景から、近年では、H.264 符号化映像の主観画質を改善することを目的として、H.264 エンコーダにおける符号化モード選択手法を最適化する研究が行われている。文献[3]では、Bピクチャにおけるダイレクトモードを適応的に用いることにより、主観画質を保ちつつ符号量を削減する手法が示されている。しかしながら、同手法における符号化モード選択はR-D最適化法の枠組みに基づいており、テキストチャ再現性に関しては考慮がなされていない。文献[4]、[5]では、主観画質の改善を目的とした符号化モード選択手法の枠組みが示されている。文献[4]では、R-D最適化法におけるコスト関数について、絵柄に応じて重み付けを行う手法が示されているが、テキストチャの再現性も含めて主観画質を改善するためのパラメータ設定に関して、具体的な方法が明示されていない。また文献[5]では、SSIM(Structure Similarity)[6]に基づく符号化モード選択手法が示されている。しかしながら、同手法は単一のMBに対する評価のみしか行っていないため、局所領域におけるテキストチャ再現性の低下を検知する

ことができず、テキストチャの再現性が失われることによる主観画質の低下を抑制することは困難である。したがって、近傍MBも含めた局所領域における主観画質という観点において、文献[3]から文献[5]による符号化性能は文献[2]と同等であるといえる。

以上の背景から、本稿ではHDTVに対する10Mbps以下の低ビットレート符号化を対象として、主観画質の改善を目的とした符号化モード選択手法の提案を行う。提案手法では、処理MBに対して複数の近傍MBの集合におけるテキストチャ再現性の低下を評価し、同低下が見込まれる領域では高いテキストチャ再現性が得られる符号化モードを選択する。さらに主観画質の評価方法としてMOS(Mean Opinion Score)を用い、提案手法の性能評価を行った。

## 2. 従来手法の課題

### 2.1. R-D 最適化法

MB 単位の符号化において、符号化に起因する符号化歪を  $D$ 、符号量を  $R$ 、割り当て符号量を  $R_c$  とする。文献[2]では、MB 単位の符号化における符号化モード選択において、割り当て符号量の制約の下で客観画質を最大化することを目的とし、式(1)の最小化問題が示されている。すなわち、式(1)の解となる符号化モードが客観画質の観点から符号化効率最大となる符号化モードであることを表している。また、式(1)にラグランジュ乗数を導入することで、式(2)のコスト関数が定義されている。式(2)は、コスト値  $J$  が最小となる符号化モードを選択することで符号化効率が最大となることを表している。

$$\min \{D\}, \text{ subject to } R < R_c \quad (1)$$

$$J = D + \lambda \times R \quad (2)$$

### 2.2. 主観画質の低下

上述のR-D最適化による符号化モード選択に起因して、動領域に特有の問題として、以下に代表されるテキストチャの再現性低下が発生する。

- イントラ符号化モードの集中による類似領域の拡大
- 符号量重視の符号化モード選択によるエッジ成分の再現性低下

### 2.2.1. イントラ符号化モードの集中による類似領域拡大

イントラ予測では、隣接する符号化済みMBの画素値を予測値として用いる。ここで、高速なカメラワークなどにより絵柄が不明瞭である映像においては、カメラワークの方向に沿って似た画素値を持つ画素が並ぶため、カメラワークの方向と予測方向が近いイントラ予測により高精度な予測が期待できる。同映像に関しては、QP値が高い低ビットレート符号化においても、符号化歪みは小さく保たれる。したがって、客観画質の観点では、同映像におけるイントラ予測は非常に効果的である。しかしながら、イントラ予測適用の結果、適用領域における画素値は近傍の画素値に近い値をとる傾向があり、入力映像のテクスチャが失われる。したがって、同領域においてイントラ符号化モードが集中した場合、類似領域の拡大によってテクスチャが単調となる。このような類似領域の拡大による弊害は動領域において検知されやすく、動きの質感が失われる結果、著しい主観画質の低下を招く。

一方で、インター符号化モードに関してはブロック単位で予測値を得るため、イントラ符号化モードよりもテクスチャが類似する可能性は低い。したがって、インター符号化モードが局所的に集中した場合でも、上述の主観画質低下は発生しない。

上記問題を実画像上で確認するため、符号化画像における局所領域の類似性に関し、イントラ符号化、インター符号化の双方について分析した。ITE HDTV 標準動画像シーケンス“Soccer action”を対象として、式(3)により局所領域の類似性  $s$  を求めた。なお、同映像は、サッカー選手およびサッカーボールの動きに追従した高速なパニングを含んでいる。図1は同映像中のあるフレームを示しており、以下では同図中の太枠で示す領域に着目する。式(3)において、 $R$  は処理MBを表し、 $T$  は処理MBの左、左上、上および右上に接する計4つのMBに相当する画素領域であって、 $t$  は  $T$  の内部に位置する任意の単位領域 (16画素×16ライン) を表す。 $t$  はMB位置に限定されず、画素単位でスライド可能とする。図2に  $R$  および  $T$  の関係を示す。

また、 $p()$  は 16画素×16ライン内の画素値であり、 $x,y$  は同領域内における画素位置を指す。図3に類似性の測定結果について一例を示す。同図は、特定のフレームを対象に画面内でイントラ符号化により類似性の増大が顕著となる箇所を抽出した上で、イントラおよびインターの各符号化画像について、類似性の大きさを色で塗り分けている。同図下段は、色と類似性の大きさの対応を示している。また、比較のため原画像に対する結果も示す。結果より、イントラ符号化MBが局所的に集中すると、インター符号化MBが集中した場

合に比べて局所領域の類似性が高くなり、テクスチャが単調になる危険性があることがわかる。

$$s = \min_{t \in T} (\sum_{x,y} (p(R, x, y) - p(t, x, y))^2) \quad (3)$$

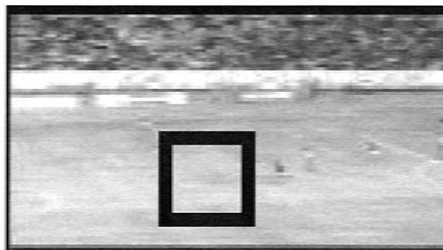


図1 対象映像

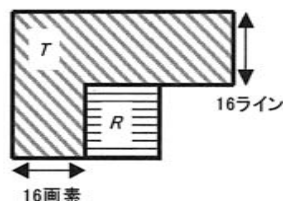
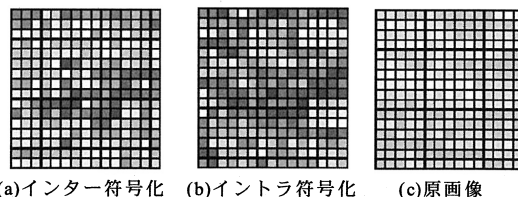


図2 R および T の関係



(a) インター符号化 (b) イントラ符号化 (c) 原画像

類似性 0 16384

図3 局所領域の類似性

以上より、従来のR-D最適化法の枠組みでは不十分であり、上記のケースにおいて高い主観画質が得られる符号化モード選択手法の枠組みが求められる。

### 2.2.2. エッジ成分の再現性低下

低ビットレート符号化でのR-D最適化法では、式(2)に示すコスト関数において符号量に大きな重み付けがされるため、符号量の小さい符号化モードが優先して選択される。これに起因する問題として、平坦領域でのエッジ成分の欠落が挙げられる。以下では、低ピッ

トレートの符号化に特化して、エッジ成分を含む MB における符号化モード選択の処理を解析する。イントラ符号化では、符号化に要するサイド情報が少ない大きなブロックサイズの符号化モードが優先して選択される。インター符号化に関しても、イントラ符号化と同様に、符号化歪みが小さい符号化モードよりもサイド情報（とくに動きベクトルが支配的）の符号量が小さいモードが優先される。したがって、従来手法に基づく符号化モード選択では、いずれの符号化モードによる予測であっても予測画像におけるエッジ成分の欠落を免れない。このとき、エッジ成分は残差信号の AC 係数として現れるものの、同係数の値は小さく、粗い量子化の適用によって除去される可能性が高い。結果、エッジ成分が失われることで主観画質の低下を招く。特に、平坦領域における動オブジェクトのエッジ成分に関して、符号化歪みが大きな値として表れない一方で動きベクトル等のサイド情報が多くなるため、符号量最小化規範と等価な符号化モード選択が行われ、エッジ成分の再現性低下が著しい。

上記を実画像上で確認する。HDTV 標準動画シーケンス“Yachting”を対象として、従来手法に基づいて符号化モードを選択したときの符号化歪みを求めた。なお当該映像は、ヨットの動きに追従した単調なカメラパンを伴っている。図4に同映像のあるフレームを示しており、以下では同図中の太枠で示す領域を対象とする。図5に、図4の太枠で示す領域における、MBごとの符号化歪みの大きさを示す。同図下段は、色と符号化歪みの大きさの対応を示している。結果より、符号化映像においてエッジ成分の再現性が低下していることがわかる。また、符号化歪みの結果からも、エッジ成分を含む MB において符号化歪みが大きくなるケースがあり、同領域においてエッジ成分の再現性が低下していることがわかる。

以上より、低ビットレートでのエッジ成分保存の観点からは、従来の R-D 最適化法の枠組みでは不十分であるといえる。

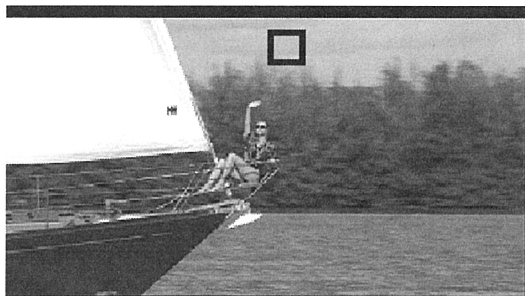


図4 対象映像

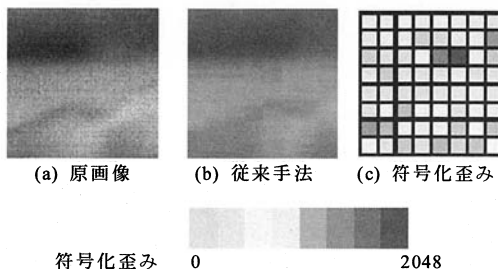


図5 エッジ成分を含む領域における符号化歪み

### 3. 提案手法

本研究で提案する符号化モード選択手法は、処理 MB 近傍の局所領域におけるテクスチャ再現性の低下を検知するものであり、R-D 最適化法の枠組みをサブセットとして持つ。すなわち、2章で示した問題を解決するために、動領域に含まれ、かつ、以下のいずれかまたは両方に該当するケースについて、提案手法に基づく符号化モード選択を適用する。

- (A) P ピクチャまたは B ピクチャにおいて、当該 MB の近傍でイントラ符号化が集中する可能性が高い
- (B) 平坦領域におけるオブジェクトのエッジ再現性が低下する可能性が高い

最初に、処理フレームのグローバル動きベクトルの大きさに基づき、当該 MB が動領域に属するかを判定する。グローバル動きベクトル  $g_v$  の大きさについて、閾値  $T_{mv}$  より大きいとき、当該 MB は動領域に属すると判定する。なお、グローバル動きベクトルを求める処理の詳細は文献[7]に従う。

結果、動領域に属する場合、以下の(a)および(b)の適応制御を行う。なお、動領域に属さない、もしくは(a)および(b)がともに適用されなかった場合は、従来のモード選択と等価な制御となる。

#### (a) イントラ符号化モードの制御

符号化映像および原画像について、処理 MB における類似性  $s$  と  $s_{org}$  を、式(3)より求める。 $s$  と  $s_{org}$  の差分の絶対値を  $D_s$  とし、 $D_s$  が閾値  $T_s$  よりも大きいとき、イントラ符号化モードの局所的な集中によるテクスチャの単調化を防ぐため、当該 MB における符号化モードの候補からイントラ符号化モードを除外する。



(b) モード選択基準の変更

処理 MB に対して  $n$  個の近傍 MB を判定領域として指定する。ここで、入力映像の画素値について、対象領域における分散  $e$  を求める。分散  $e$  が閾値  $T_e$  以下である場合、平坦領域における動オブジェクトのエッジ成分再現性低下を防ぐため、処理 MB におけるモード選択は予測誤差を最小化する規範とする。

4. 評価実験

提案手法について主観評価実験を行った。エンコーダとして JM10.1 をベースとし、3 章に基づく手法を実装した。本実験で用いた提案手法における制御パラメータを式(4)に示す。

$$\begin{cases} T_{mv} = 4 \\ T_s = 8192 \\ T_e = 50 \end{cases} \quad (4)$$

評価用映像として、HDTV 標準動画シーケンス "Yachting", "Whale show", "Soccer action" を用いた。各素材ともシーケンス長は 300 フレームである。各素材の映像特徴を表 1 に示す。符号化レートは 8Mbps, 10Mbps, 13Mbps であり、符号化条件を表 2 に示す。

表 1 素材の特徴

素材名	特徴
(A) Yachting	- ヨットの動きに追隨したカメラワーク(カメラパン) - 単調な動き
(B) Whale show	- シャチの動きに追隨したカメラワーク(カメラパン) - 高速で細かな動きを伴う
(C) Soccer action	- サッカー選手の動きに追隨したカメラワーク(カメラパン) - 高速な動きが含まれる - 観客の描画に関して詳細なテキストが含まれる

表 2 符号化条件

設定項目	設定値
プロファイル	High Profile
レベル	4.0
符号化方式	CABAC
動き予測	UMHexagon Search
GOP 構成	IBBPBBPBBPBBPBB
参照枚数	2
動きベクトル探索範囲	$\pm 64 \times \pm 64$
符号化構造	MBAFF
DCT 変換サイズ	$4 \times 4$ または $8 \times 8$

提案手法による符号化画像について主観評価実験を行った結果を表 3 に示す。また比較のため従来方式として JM10.1 による結果も示す。なお、主観評価実験における測定条件を表 4 に示す。結果より、本研究が想定する 10Mbps 以下の低レート符号化において、従来手法に対して主観画質の改善効果があることがわかる。特に、素材(A)において、従来手法に対して MOS が最大 1.1 程度向上していることがわかる。

また、表 3 におけるビットレートの変化に対する MOS の推移より、提案手法における 8Mbps の結果を従来手法で得るためにはおよそ 18Mbps 程度のビットレートが必要であると見込まれる。このとき、従来手法に対する提案手法の符号量削減量は、およそ 55% 程度であることとなる。

表 3 MOS の結果

素材	8Mbps	10Mbps	13Mbps	
(A)	従来(JM10.1)	3.08	3.27	3.58
	提案手法	4.17	4.25	4.25
(B)	従来(JM10.1)	3.67	3.67	3.83
	提案手法	3.67	3.83	3.83
(C)	従来(JM10.1)	3.25	3.40	3.75
	提案手法	3.58	3.60	3.67

表 4 主観評価実験の測定条件

測定距離	3H
ディスプレイ	24 inch CRT
被験者	15(非専門家)
測定手法	一重刺激法 (ITU-R BT.500-11 [8]準拠)

提案手法により主観画質の改善が見られた素材(A)と素材(C)において、効果分析を行う。まず素材(A)については、従来手法で主観画質の低下が見られた領域内の半数程度の MB で 3 章(b)の制御が適用され、有効性が読み取れる。上記制御が適用された領域での改善効果を分析するため、各 MB における符号化歪の大きさを確認した。図 6 に従来手法および提案手法について、原画像との差分二乗和を MB ごとに集計した結果を示す。結果より、提案手法の効果として、輪郭におけるエッジ成分が安定的に保持された結果、主観画質の向上に寄与しているといえる。

次に素材(C)に関して、提案手法により主観画質の改善が得られた領域では、3 章(a)の制御が多く適用されている。同制御が適用された領域における類似性について確認する。図 7 に同領域における類似性を示す。取得条件は 2 章に従うこととする。結果より、提案手法により局所領域内の類似性が抑制される方向にあり、これにより類似領域の拡大を抑制し、主観画質が改善していることが読み取れる。

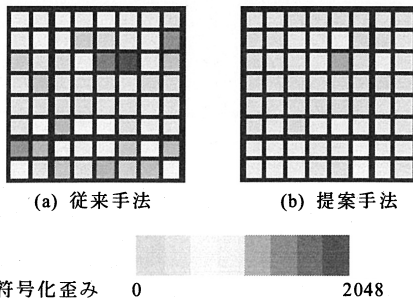


図6 復号画像における符号化歪み(素材(A))

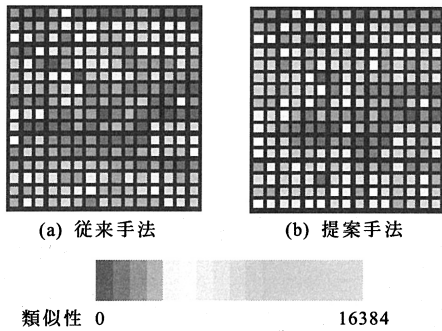


図7 類似性(素材(C))

3章で示したとおり、処理MBにエッジ成分が含まれるとき、提案手法のモード選択手法では予測誤差が最小化される。本稿で扱う低ビットレート符号化において、符号化歪みとMC予測誤差の大小関係はほぼ等価である。したがって、予測誤差を最小化することは、R-D最適化法において $\lambda=0$ とすること、ほぼ等価である。実験結果からも、両者で同様の結果が得られることを確認している。

より高ビットレートでの符号化において提案手法を適用した場合、表3の結果より、素材(C)に対する13Mbpsの符号化を除き、従来手法に対して同等以上の結果が得られている。従来手法で良い結果が得られている場合については、イントラ符号化の適用制限が過度に適用され、符号化効率が低下したためであると考えられる。より高ビットレートでの符号化における適応制御の最適化については、今後の課題である。

最後に、提案手法におけるモード選択処理に要する計算量を評価した。1GOP(15フレーム)の符号化処理に要した処理時間を表5に示す。処理時間は従来方式としてJMによる処理時間を1.0として換算した値を用いた。実験環境は、汎用的なPC環境であり、CPU3.2GHz、

メモリ2GBである。結果より、提案手法によるモード選択手法に要する時間の増加は、約0.5%程度である。したがって、提案手法による処理時間の増加は無視できる程度であることがわかる。

表5 処理量(従来手法を1.0として換算)

素材	処理量
(A)	1.008
(B)	1.005
(C)	1.006

## 5. まとめ

本研究では、HDTVを対象としたH.264低ビットレート符号化において、主観画質を改善するモード選択手法を提案した。提案手法は、局所領域における主観画質低下要因を検知する新たな評価尺度を備え、これに基づいたモード選択を行うことでテクスチャの再現性を追及したものである。主観評価実験より、提案手法の適用によって、従来のモード選択方式に対する主観画質の改善効果があることを確認した。

謝辞 日頃御指導いただく株式会社KDDI研究所代表取締役所長 秋葉重幸博士に感謝の意を表す。本研究は独立行政法人情報通信研究機構による委託研究「ソフトウェア符号化技術に関する研究開発」として実施したものである。

## 文 献

- [1] Karsten Suhring, H.264/AVC Software Coordination, <http://ip.hhi.de/index.htm>
- [2] Optimization for Video Compression", IEEE Signal Processing Magazine, pp.74-90, Nov. 1998.
- [3] Adriana Dumitras and Barry G. Haskell, "Enhancement of Direct Mode Selection in B Pictures for Bit Rate Reduction of Compressed Video Sequences", ICIP2003, III-825-8, Sep. 2003.
- [4] Koohyar Minoos and Trung Q. Nguyen, "Perceptual Video Coding with H.264", Signal, Systems and Computers 2005, pp.741-745, Oct. 2005.
- [5] Zhi-Yi Mai, Chun-Ling Yang, Kai-Zhi Kuang and Lai-Man Po, "A Novel Motion Estimation Method Based on Structural Similarity for H.264 Inter Prediction", ICASSP2006, II-913 - II-916, May 2006.
- [6] Z. Wang, A. C. Bovik, H. R. Sheikh, and E.P. Simoncelli, "Image quality assessment from error visibility to structural similarity," IEEE trans. Image Processing, vol. 13, no.4, pp. 600-612, Apr. 2004.
- [7] F. Dufaux and J.Konrad, "Efficient, robust, and fast global motion estimation for video coding," IEEE Transactions on Image Processing, Vol.9, No.3, pp.497-501, Mar 2000.
- [8] ITU-R BT.500-11, "Methodology for the subjective assessment of the quality of television pictures", 2002.