

携帯電話によるビデオ配信向け 高機能コンテンツ制作システム

滝嶋 康弘 米山 暁夫 宮地 悟史 柳原 広昌 中島 康之 和田 正裕

株式会社 KDDI 研究所
〒356-8502 埼玉県上福岡市大原 2-1-15

E-mail: {takisima, yoneyama, miyaji, yanagihara,
nakajima, wada}@kddilabs.jp

あらまし 携帯電話に対するビデオ配信用コンテンツ制作システムの設計法を提案する。モバイルマルチメディアが注目される中、そのコンテンツ制作は従来のインターネット配信向け制作環境などとは異なる要求条件を有する。受信端末環境、伝送ネットワーク環境、制作環境における低レート伝送、素材時間短縮、補助情報・テロップ挿入、一素材多利用等の条件を考慮し、低レート AV 符号化、高機能 AV 編集、軽量テロップフォーマット、高速フォーマット変換等の特徴を持つ制作システムを提案する。本設計に基づくソフトウェアシステムの開発についても報告する。

キーワード ビデオ配信、モバイル通信、コンテンツ制作、コンテンツフォーマット、テロップテキスト

Functional Content-authoring System for New Generation Mobile Video Communications

Yasuhiro TAKISHIMA, Akio YONEYAMA, Satoshi MIYAJI,
Hiromasa YANAGIHARA, Yasuyuki NAKAJIMA and Masahiro WADA

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

E-mail: {takisima, yoneyama, miyaji, yanagihara,
nakajima, wada}@kddilabs.jp

Abstract This paper proposes the comprehensive architecture of a novel content-authoring system specific for mobile communication services. Considering unique requirements to mobile communications such as receiver property, transmission environment, and content-creating conditions, the authoring-system should be equipped with a high-quality low-bit-rate video encoder, a functional AV editor, a supportive caption inserter, a fast format converter, and a user-friendly previewer. Software system based on the proposed architecture is also introduced.

Key words Video Content Delivery, Mobile Communications, Content Authoring, Content Format, Auxiliary Captions

1. はじめに

携帯電話によるマルチメディアサービスが注目されている。特に通信速度の向上とともに、今後いわゆるリッチコンテンツと呼ばれるマルチメディアコンテンツ、特にビデオコンテンツの配信が発展するものと期待されている。こうしたコンテンツの流通のためには、これらを効率的にかつ高品質で制作するオーサリングシステムが不可欠である¹⁾。しかし、携帯電話でのコンテンツの受信や視聴は他のネットワーク・機器とは、その動作環境が大きく異なるため、コンテンツ制作にも大きな影響を与える。こうした環境の相違を十分に考慮した制作システムは従来存在していない。本稿では、携帯電話での配信に特化したビデオコンテンツ制作システムについて、その設計法を提案する。

2. 機能要求条件とシステム設計

2-1 機能要求条件

インターネット配信などを目的としたコンテンツ制作環境とは異なる条件が必要となる。例えば、受信端末環境を考慮した場合、表示画面サイズに適したビデオフォーマット、記憶容量の制約からデータサイズの小ささ、オーディオ以外のビデオ補助用付随メディア、などが必要となる。また伝送環境の観点からは、コンテンツ長の短縮や高いメディア符号化効率が要求される。一方制作環境を考慮した場合、一般的な編集機能に加えて複数コンテンツの高速連続制作や一素材多利用などの機能が必要である。

2-2 システム設計

上記条件を考慮して、制作システムの基本機能および相互連携による処理フローを図1に提案する。AV信号の直接入力で作成を行う Encoder、コンテンツ形式変換を行う Converter、基本カット編集および先頭・末尾の部分削除を行う Editor/Trimmer、補助メディアとしてテロップテキストを作成・挿入する Telopper、作成コンテンツおよび処理パラメータ確認用の Previewer から構成される。

Encoder は外部 AV 信号を、画像サイズ・データサイズ・符号化効率などの要求条件に合致した提案形式のファイルへと出力する。コンテンツの一素材多利用を実現するため、Converter は非圧縮形式、あるいは既制作の別形式ファイルを入力とし、高速変換処理により提案形式で出力する。連続制作機能も有する。同様に非圧縮形式や別形式ファイルに対

して、コンテンツ長の短縮が必要な場合には、Editor/Trimmer により、入力が圧縮形式の場合でも画質劣化を抑制しながら、カット編集を行う。オーディオメディアが必ずしも効果的に利用できない受信端末環境に対して、補助メディアとしてテロップテキストの付加を提案し、その作成・挿入を Telopper が行う。テロップテキストは、ビデオコンテンツとの同期が必要なため、最終段階で作成する。

制作コンテンツの画質、内容、処理パラメータを確認し、再調整を行うために Previewer がこれらを表示する。受信端末ごとの表示機能の差異を考慮し、想定受信端末に応じた模擬表示機能を有する。

本制作システムの機能仕様を表1に示す。提案の構成では、機能部位を分割しているため、全体として機能上のスケラビリティを有することができる。以下、本制作システムにおける各機能やデータフォーマットの詳細について提案する。

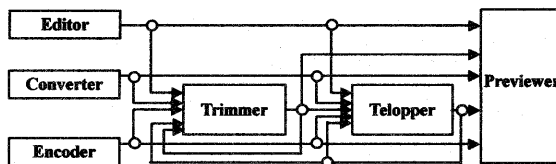


図1 機能構成と処理フロー

表1 システム機能仕様

機能	仕様
Video 符号化	MPEG-4 Simple Profile
Audio 符号化	Speech Enc (Monaural) / Music Enc (Stereo)
MUX 形式	MP4
補助メディア	テロップテキスト
制作機能	編集・トリム、変換、バッチ処理
確認機能	パラメータ画質確認プレビューワー

3. コンテンツフォーマット

3-1 コンテンツフォーマット概要

携帯電話でのビデオ配信では、端末の限られた CPU リソースや与えられた伝送ビットレートの範囲で、要求される機能を実現するため、伝送プロトコルや多重化フォーマット等について検討する必要がある²⁾。携帯電話でのビデオ配信に適したコンテンツフォーマットに関して、要求条件の検討を行い、実現方法について提案する。

3-2 要求条件

携帯電話端末のリソースを有効活用する、不要な処理はできるだけ行わない、という観点から、以下のような機能および実装面での要求条件が挙げられる。

- ・ 伝送プロトコルは WAP 対応 HTTP/TCP/IP
- ・ 多重分離、復号で CPU パワーを消費しないこと
- ・ テロップ等の関連データを格納できること
- ・ 長時間コンテンツにも対応できること
- ・ 受信しながらの再生が行えること

これら要求条件を満足するためのコンテンツフォーマットについて、以下に詳細を提案する。

3-3 フォーマットの詳細

まず伝送プロトコルについては、すでにモバイル端末でインターネットにアクセスする際に使用されている WAP Object Download 準拠とする。これにより、WWW ブラウザで使用しているプロトコルスタックと共通化できるため、内部リソースの有効利用ができ、さらにアプリケーションレベルでのサーバ/端末制御を必要としないため、既存サーバの有効利用や端末内のクライアントソフトウェアの小型化に貢献できる。

一方、アプリケーションレベルでのサーバ/端末制御を行わないため、多くの機能がファイルフォーマットに依存することとなる。したがって要求条件を満足するためのファイルフォーマットを詳細に検討する必要がある。ここでは、国際標準性を考慮して、MPEG-4 System^[3]で規定されている MP4 (ISO Media File Format から派生)^[4]をベースとする。MP4 フォーマットは、ファイル全体情報を記述するヘッダ、各メディアデータ、ならびにユーザ定義情報を構成要素とし、これらの構成にはかなりの自由度がある。今回、要求条件を考慮し、メディア同期や端末でのファイルハンドリングを容易にするため、各要素を単一ファイル上に多重化配置することとし、これを図2に示す。

図2に示す通り、メディアデータはチャンクと呼ばれる一定の単位(数フレーム)ごとに多重化される。ここでチャンクサイズは、一度に端末に読み込まれるメモリサイズおよび、かつ多重分離処理の頻度(CPU 負荷)に影響を与えるため、最適なサイズとしなければならない。

また、メディアデータの符号化方式については、これまで携帯端末での実装検討が十分に行われてき

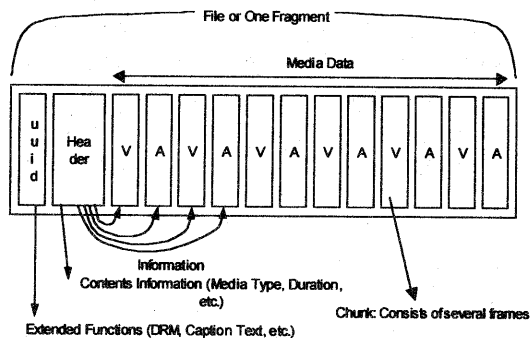


図2 コンテンツフォーマットの構成

た MPEG-4 Visual SP@L1, すでに音声通話に使用されているスピーチコーデック, ならびに楽音用コーデックをサポートし, 幅広いジャンルのコンテンツの作成に対応する。

さらに受信しながらの再生(ストリーミング)に対応するため、ヘッダ情報や、テロップ・著作権などのコンテンツ関連情報をファイル先頭に配置する。その際、長時間コンテンツでは、ヘッダ情報が大きくなるため、ストリーミング受信においては、ヘッダの受信が大きな初期オーバーヘッドとなる。また一度に大きなヘッダ情報を端末内に格納しなければならないという問題が生じる。これに対処するため、MP4で規定されている Fragment 構造を利用し、分割格納することとする。

このようにして作成されたコンテンツは、サーバ上に置かれ、Web ページを経由して携帯端末からダウンロードされ、端末上で再生されることとなる。

4. コンテンツフォーマット変換機能

4-1 コンテンツフォーマット変換機能の概要

大量のコンテンツを効率よく制作するためには、既存コンテンツの有効活用が重要な課題である^[5]。ここでは、携帯電話での配信に特化したビデオコンテンツ制作システムにおける、携帯電話向けフォーマットへのコンテンツ変換方式について提案する。

4-2 コンテンツフォーマット変換要求条件

PCソフトウェアによるコンテンツ編集、制作環境が普及しており、その場合に用いられる素材のフォーマットは AVI や MPEG が主である。そのため

これらのコンテンツを受信端末環境に適した画像サイズ、データサイズ、フォーマットに変換することが必要となる。またこの変換の際に、複数の素材から同一の出力設定でコンテンツをまとめて作成したり、同一の素材から複数の受信端末環境に適したコンテンツをまとめて作成するという機能も必要となる。

4.3 コンテンツフォーマット変換方式

入力されたコンテンツは、デマルチプレクス、A・Vデコード、A・Vエンコード、マルチプレクスの4ステップで変換を行う。ビデオについては復号、リサイズ処理を行った後、MPEG-4方式で所望の画質または所望のフレームレートでの符号化を行う。また出力コンテンツのフレームレートが入力コンテンツのフレームレートに対して低い場合には、コンテンツ生成に必要な画像以外の画像のデコードは行わないことで処理の高速化を図る。

また、制作したモバイル用ビデオコンテンツから再符号化を行わずにトリミング編集を実現するためには制限を伴う。そこで、モバイル用ビデオコンテンツ作成時にI-VOP挿入間隔を時間単位で設定可能とし、設定されたI-VOP挿入間隔と、解析した入力ビデオコンテンツのシーン切替り位置からI-VOPの挿入位置をシーンチェンジに同期させた。これによりシーンの切替り点でのコンテンツのトリミングを容易に実現することができる。

また、オーディオについても所望の設定でのモノラル/ステレオ符号化を行い、生成されたA・VデータはMP4フォーマットとして多重化される。

図3に変換ソフトウェアのGUIを、表2に機能仕様を示す。変換を行う複数の設定についてはリスト形式により表示する。変換速度はPentiumIII 700MHzのPCで画像サイズ320x240,30fps,30秒のMPEG-1コンテンツから画像サイズ176x144,10fpsのコンテンツを作成するのに約7秒であった。

このようにして生成したオーディオ、ビデオコンテンツに対しては、補助メディアとしてテロップテキストの挿入を行うことが可能である。

5. 編集機能

5-1 編集機能の概要

国際標準規格MPEG-4 System(ISO/IEC 14496-1)で策定されたMP4ファイルフォーマットをもとに携帯電話向けのビデオ配信サービスのコンテンツフォーマットを規

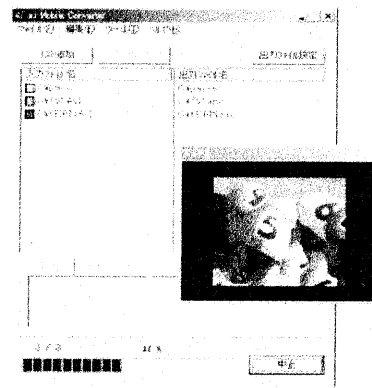


図3 コンバータ 操作画面

表2 コンバータ機能仕様

機能	仕様
入力フォーマット	AVI, MPEG
Video 符号化	MPEG-4 Visual Simple Profile
Video サイズ	128x96, 176x144
Audio 符号化	Speech Enc (Monaural) / Music Enc (Stereo)
MUX 形式	MP4

定した。特に時間長についての制約が強いことを考慮すると、コンテンツ長の微調整のために、先頭や末尾の切り取り(トリミング)機能が重要となる^[6]。ここでは、編集方式のひとつとしてのトリミング機能について提案する。

5-2 トリミングによる同期ずれの問題

MPEG-4 Systemの蓄積用フォーマットとしてMP4ファイルフォーマットが規定されている。今回はMP4上でのトリミング編集を検討するため、エレメンタリーストリーム(ES)の編集は考慮しない。そのため切り出し開始点はランダムアクセスポイント(RAP)とする。ここで問題となるのは複数のESのRAPが必ずしも一致せず、同期再生にずれが生ずる点である。MP4の仕様上、復号開始の初期時刻は0と決められているため、全てのESが完全に同時刻から再生される場合を除き、デコード開始時刻をRAPのずれだけシフトする必要がある。しかし、これによりESが同時に再生開始できなくなる。また、シフト情報を格納する領域はオプションであるため、この情報を利用しないデコーダの存在が懸念される。更に、再編集によってトリミングが繰り返されると同期ずれが蓄積される問題が憂慮される。よって、MP4ファイルのトリミング編集においてESの同期ずれ抑止する方法の確立が必要となる。

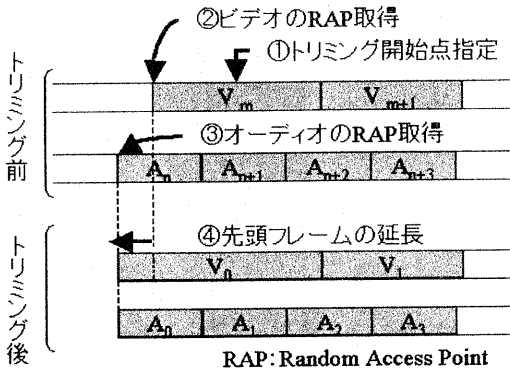


図4 同期あわせのための第一フレームの延長

5-3 同期ずれの補償方式

トリミングの開始時刻が入力されると、開始時刻を越えない最大の復号時刻を持つ RAP を取得する。但し、同期ずれを最小に抑えるため、複数の ES の中で RAP 間隔が最も広いビデオ ES を対象とする。オーディオの切り出し点はビデオの RAP を超えない最大の RAP とする。このとき、最大でオーディオ 1 アクセスユニット (AU) の再生時間のずれが生じるので、ビデオの先頭フレームと次フレームの再生間隔にずれを埋め込むことで同期ずれを解消し、2 フレーム目以降の同期再生を維持することを提案する。オーディオを常にビデオより過去の RAP から開始するため、ビデオの 1 フレーム目と 2 フレーム目の間隔は RAP の再生時刻の差分だけ広がる。これはビデオのビットレートが低下することと等価で、一定量のデータが得られるまでデコードを開始しない VBV バッファモデルには影響を与えない。また、MP4 として格納される時刻情報は復号時刻の差分値を格納するので、ES の再生時刻全体がシフトしても変更点は一ヶ所に抑えられる。一方、再編集によるずれの蓄積は 2 フレーム以降のストリーム構成が編集前と変わらないため、最大の同期ずれは初回のトリミングによるオーディオ 1AU 分のみに抑えられる。1 フレームと 2 フレーム間に切り取り開始点を指定した場合も RAP がともに 0 であることから、同期ずれは生じない。

6. テロップテキスト機能

6-1 テロップテキスト機能概要

テロップ情報は、オーディオビジュアルコンテンツの表現力を飛躍的に高める効果がある。特に、オーディオメディアが必ずしも再生できない環境に受

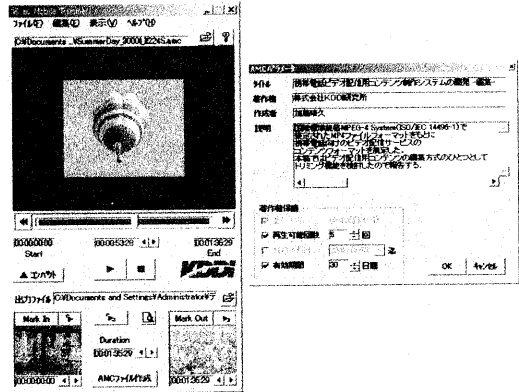


図5 制作コンテンツ編集用 Trimmer

信端末が置かれている場合、ビデオの補助メディアとしての意義は大きい。ただし、符号化ビットレートや携帯電話の表示解像度には制約があるため、ビデオに重畳して符号化した場合、可読性が悪くなる恐れがある[7]。そこで、テロップ情報をビデオやオーディオとは別のメディアとして扱うためのフォーマットを規定し、それを制作する環境を構築する。

6-2 テロップテキスト機能要求条件

最も基本的な機能は、ビデオと同期して文字列を表示することである。そのため、文字列だけでなく、表示・終了時刻を規定できる必要がある。ついで、表現力を高めるために様々な装飾的表示効果 (表 3) が必要となる。

表 3 テロップ表示効果

時間的挙動	
スクロール	文字列の移動
ウェイト	表示の一時待機 (用例: 一文字ずつ時間を空けて表示する)
ブリンク	文字の点滅
ワイブ	文字色の変化 (用例: カラオケ)
文字修飾	
アンダーライン	下線
フォント色	文字の色
反転	文字色と背景色の入れ替え
リンク	
url	Web、電話、メールへのリンク

6-3 テロップテキストフォーマット

テロップ情報の表現には、XML ベースのフォーマットを用いる。MP4 の atom ベースの表現と比較して、将来の拡張性の高さ、テロップ情報の編集のしやすさなどの利点がある。テロップ情報の内部構造を以下に示す。表現形式は SMIL ライクであるが(図6)、携帯電話でのテロップ表示エリアは限られていることから、表示位置などレイアウトに関する表現は省略している。なお、MP4 ファイルへの格納にあたっては、ユーザプライベートデータのための uuid atom を用いている。

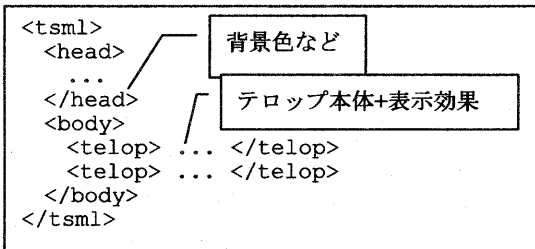


図6 テロップテキスト表現形式

6-4 テロップ・アプリケーション

本稿で提案したテロップ情報を作成するためのツールであるテロップの概略は以下の通りである。

テロップ情報はビデオコンテンツのシーンごとに IN 点、OUT 点を設定し、そのシーンに対してテキストおよび表示効果を定義するという手順で作成する。このため、テロップは図7に示す GUI を持っており、ビデオを参照しながら IN/OUT 点を容易に設定できるようになっている。また、テキストの挿入や表示効果の定義についても、直感的な操作で行うことが可能なよう設計されている。

7 おわりに

作成されたコンテンツは、受信端末での表示を模擬した Previewer で再生・表示可能とする(図8)。受信端末種別に応じた表示サイズ、表示画面のレイアウト等が選択できるほか、符号化パラメータ等の参照機能を有し、コンテンツの再制作が容易となる。

また、機能単位の構成のため、多機能編集システム等、他の制作機能との組み合わせが容易である。システムスケーラビリティを考慮したワークフローの一層の効率化に関して今後検討する。

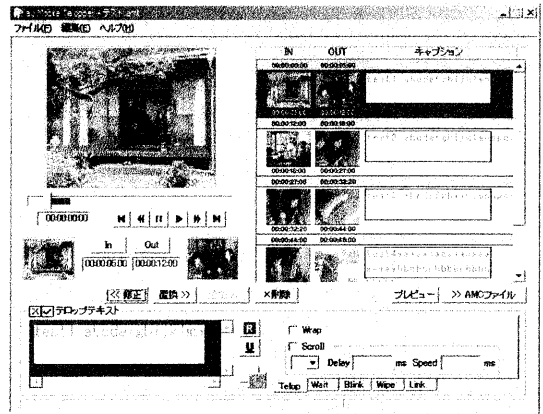


図7 テロップ・アプリケーション GUI

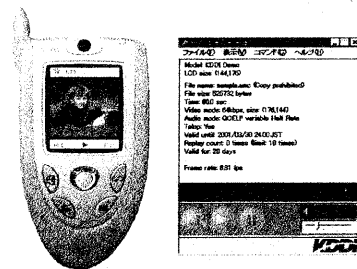


図8 制作コンテンツ確認用 Previewer

文 献

- [1] 滝嶋、米山、宮地、柳原、中島、和田：「携帯電話ビデオ配信用コンテンツ制作システムの開発 -基本設計-」、映像情報メディア学会 2001 年冬季大会
- [2] 宮地、加藤、高木、柳原、中島、和田：「携帯電話ビデオ配信用コンテンツ制作システムの開発 -コンテンツフォーマット-」、映像情報メディア学会 2001 年冬季大会
- [3] ISO/IEC: "ISO Media File format specification CD," Jul. 2001
- [4] ISO/IEC: "MP4 version 2 CD," Jul. 2001
- [5] 米山、柳原、宮地、中島、和田、「携帯電話ビデオ配信用コンテンツ制作システムの開発 -フォーマット変換方式-」、映像情報メディア学会 2001 年冬季大会
- [6] 加藤、米山、柳原、中島、和田、「携帯電話ビデオ配信用コンテンツ制作システムの開発 -編集方式-」、映像情報メディア学会 2001 年冬季大会
- [7] 酒澤、高木、宮地、加藤、中島、和田：「携帯電話ビデオ配信用コンテンツ制作システムの開発 -テロップテキストフォーマット-」、映像情報メディア学会 2001 年冬季大会