

パーソナライズ映像配信システムの試作

山田 真理子[†] 鈴木 敏明[†]

[†] (株)日立製作所 中央研究所 〒185-8601 東京都国分寺市東恋ヶ窪 1-280
E-mail: [†] {mariko.yamada.hs, toshiaki.suzuki.cs}@hitachi.com

あらまし ビデオサーバより配信される複数の視点映像の中から、視聴者が要求したキーワードに対応する映像を、ゲートウェイにて切換えて配信するパーソナライズ映像配信システムを提案する。本稿では、ビデオサーバ及びゲートウェイの処理について検討し、パーソナライズ映像選択配信処理を高速に行う方式を提案する。提案方式では、ビデオサーバから配信する映像データの伝送パケットのヘッダに、映像の切換え可能点を示す識別子を付与し、ゲートウェイは、識別子を基に各視点映像を切換え、パーソナライズ映像の配信を行う。提案方式の試作を行い、ゲートウェイにおけるパーソナライズ映像配信性能を評価した。その結果、1台のゲートウェイにて90ユーザに配信が可能であることを明らかにした。

キーワード パーソナライズ映像、多視点映像、ストリーム切換え、MPEG、GOP

Prototype of Personalized Video Distribution System

Mariko YAMADA[†] and Toshiaki SUZUKI[†]

[†] Hitachi, Ltd., Central Research Laboratory

1-280 Higashi-koigakubo, Kokubunji-shi, Tokyo, 185-8601 Japan
E-mail: [†] {mariko.yamada.hs, toshiaki.suzuki.cs}@hitachi.com

Abstract We propose a personalized video distribution system. In this system a gateway transmits a personalized video stream, which is selected from multi-angle video streams based on comparisons between preference registered by audience and meta-data of each video stream. We studied methods of selecting a proper video stream from multi-angle video streams and transmission process at a high speed. In these methods, video servers give an identifier to the header of the transmission packet of video data. The identifier shows a possible point to change video stream. The gateway selects the video stream according to the identifier and transmits it. We implemented a prototype of this system and evaluated it. It was clarified that 90 personalized video streams were transmitted by one gateway.

Keyword Personalized video stream, multi-angle video stream, stream change, MPEG, GOP

1. 緒論

FTTH などのブロードバンド環境の整備に伴い、ネットワークを利用した映像配信が普及しつつある。現在では、インターネットサービスを提供する事業者の多くが映像配信サービスを提供している。映像配信事業者の中には、複数の視点から撮影した映像を同時に配信するサービスを提供する事業者もある[1]。さらにネットワークが高速化することにより、付加価値のある映像配信サービスが可能となる。例えば、GPON(Gigabit capable passive optical network)を用いた映像配信では、100チャンネルのハイビジョン映像の同時配信が可能である[2]。また、視聴者が任意の視点を自由に選択できる自由視点映像配信

技術の研究も行われている[3]。

このような映像配信技術の進化により、複数のチャンネルから映像を選択したり、任意の視点から映像を選択することが可能となる。しかしながら、提供されるチャンネルや視点映像の増加に伴い、視聴者が視聴を希望する映像を自ら探し出すことが困難となる。

そこで、本研究では、ネットワーク上で配信される複数の視点映像から、視聴者の要求に応じた視点映像を随時選択し、視聴者ごとにカスタマイズした映像を配信するパーソナライズ映像配信システムについて検討する。本システムにより、視聴者は簡単な入力操作だけで、希望する映像を視聴することが

可能となる。2 章では、システム構成及び、パーソナライズ映像を生成するゲートウェイの課題について検討する。3 章では、実装方法について検討し、4 章では、試作したシステムの評価を行う。

2. 提案システムの概要と課題

2.1. システム概要

パーソナライズ映像配信システムは、ビデオカメラで撮影した各視点映像をユーザに配信するとともに、それらの各視点映像の中からユーザの要求するキーワードに対応する映像を選択して配信するシステムである。図 1 にシステム構成を示す。本システムは、センササーバ、イベントサーバ、ビデオカメラ、ビデオサーバ、ゲートウェイ及びクライアントから構成する。各サーバ及びクライアントは IP ネットワークを利用して通信する。

センササーバは、各センサが検出した RFID(Radio Frequency Identification)の情報をイベントサーバに通知する。イベントサーバは、RFID とそれに対応する被写体の情報(例えば、スポーツ選手や俳優の名前等)を管理しており、センササーバから通知された RFID の情報から、ビデオカメラで撮影されている映像のキーワードを特定する。特定したキーワードをビデオサーバに通知する。ビデオサーバは、ビデオカメラで撮影した映像をマルチキャストで配信する。映像配信プロトコルは UDP(User Datagram Protocol)及び RTP(Real-time Transport Protocol)を用いる。また、イベントサーバから通知されたキーワードを映像のメタデータとしてマルチキャストで配信する。ゲートウェイは、各ビデオサーバから配信される視点の映像をマルチキャスト配信する。また、クライアントから通知されたキーワードと映像のメタデータとを比較し、各視点の映像から適切な映像を選択する。選択した映像を随時切換えながら、1 つのストリームに連結してユニキャストで配信する。各視点映像を連結して配信する映像を本システムではパーソナライズ映像と定義する。

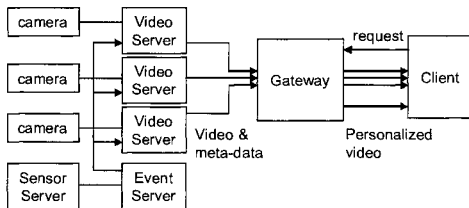


図 1 パーソナライズ映像配信システム

図 2 を用いてパーソナライズ映像の生成方法を説明する。各ビデオサーバは視点 1 から 3 の映像を配信している。視点 1 映像は、時刻 t_0-t_1 間に被写体 A、時刻 t_1-t_3 間に被写体 B、時刻 t_3-t_4 間に被写体 C を映している。視点 2 映像は、時刻 t_0-t_1 間に被写体 B、時刻 t_1-t_2 間に被写体 A、時刻 t_2-t_4 間に被写体 B を映している。視点 3 映像は、時刻 t_0-t_2 間に被写体 C、時刻 t_2-t_4 間に被写体 A を映している。パーソナライズ映像を視聴するユーザは、視聴する映像のキーワードとして被写体 A を登録している。ゲートウェイでは、A を被写体とする視点の映像を切換えてユーザに配信する。すなわち、時刻 t_0-t_1 間に視点 1 の映像、時刻 t_1-t_2 間に視点 2 の映像、時刻 t_2-t_4 間に視点 3 の映像をパーソナライズ映像としてユーザに配信する。

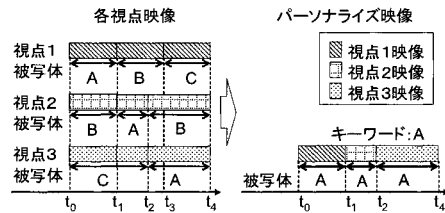


図 2 パーソナライズ映像の生成

2.2. パーソナライズ映像作成時の制約

本システムでは、映像の符号化方式として広く普及している MPEG2 を用いた。MPEG2 では、フレーム間の類似性を利用して符号化しており、符号化時に参照するデータの違いから、I、P、B の 3 種類のフレームが存在する。I フレームは、前後のフレームを参照せず単独で符号化するフレームである。P フレームは、それ以前のフレームを参照して符号化するフレームである。B フレームは、前後のフレームを参照して符号化するフレームである。これらのフレームを効率的に処理するため、I フレームから始まる一連のフレームをまとめて管理している。これらフレーム群は、GOP(Group of Pictures)と呼ばれる。

前後のフレームを参照して符号化を行う場合、任意のフレームで映像を連結すると、参照するフレームが存在せず、映像を復号できない場合がある。例えば、転送順で B フレームの直前で連結すると、転送を開始する映像は、B フレームの復号時に参照する I または P フレームが欠落しているため、フレームを復号することができない。そのため、復号に必要なフレームが欠落しないように映像の切換点を決

定する必要がある。図 3 に映像の連結開始・終了点を示す。映像の連結終了点となり得るのは、復号時に終了点以後に転送されるフレーム情報を必要としないフレーム境界であるので、I または P フレームの直前である。連結開始点となり得るのは、復号時に開始点以前に転送されたフレームの情報を必要としないフレームの境界であるので、I フレームの直前である。GOP は I フレームから次の I フレームまでの一連のフレーム群であるため、GOP 境界は常に映像の連結開始終了点になり得る。ゲートウェイでは、この制約を満たす任意の時点で映像を連結する必要がある。

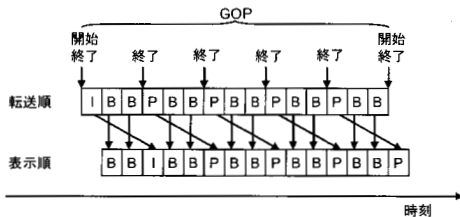


図 3 映像連結開始・終了点

2.3. 本システムの課題

本システムにおけるゲートウェイの課題について検討した。

(1) 切替点の検出課題

映像連結時の制約を満たす切替点は、映像データに含まれる開始コードの検出により判定することができる。GOP の先頭を示す GOP 開始コードは 0x000001B8、フレームの先頭を示すピクチャ開始コードは 0x00000100 である。また、フレームの種類は、ピクチャ開始コードに続く、ピクチャコーディングタイプにより判定できる。しかし、GOP 及びフレームのデータ量はそれぞれ異なるため、ゲートウェイでは、すべての映像データについて開始コードの検出処理を行う必要がある。さらに、ゲートウェイでは、これらの処理をリアルタイムに行う必要があり、切替え点検出の処理負荷が高い。

(2) 切替ストリーム間の同期課題

各視点の映像は、IP ネットワークを利用して、各ビデオサーバから非同期に配信されるため、切替点となる映像データが含まれたパケットがゲートウェイに到達するタイミングが異なる。ゲートウェイからクライアントへの映像配信が断続的に行われると、クライアントのバッファにアンダフローやオーバフローが発生し、映像品質が低下する。これを避けるために、ゲートウェイからパーソナライズ映像を配

信する際は、映像連結時の制約を満たしつつ、送信間隔を一定にする必要がある。

また、同期がとれていないストリーム間では、PTS(Presentation Time Stamp)や PCR(Program Clock reference)等のタイムスタンプ情報がそれぞれ異なる。これらを連結して1つのストリームにする場合、タイムスタンプ情報を適切に変換する必要がある。

3. 実装方法の検討

前節で検討した本システムの課題を解決する実装方法を検討した。

3.1. 検出処理低減方式の検討

ゲートウェイにおける切替点の検出処理を低減するために、ビデオサーバ及びゲートウェイの処理について検討した。

ビデオサーバは、ビデオカメラで撮影した映像の圧縮符号化及び RTP パケットの生成を行い、各視点映像を配信する。

ビデオサーバとビデオカメラとは1対1に接続している。一方、ゲートウェイは複数のビデオサーバ及びクライアントと接続しており、接続するビデオサーバ及びクライアントの増加に伴って、処理するストリームの数が増加する。ビデオサーバで扱うストリームの数は、ゲートウェイに比べて少ないため、切替点の検出はビデオサーバで行い、その結果をゲートウェイに通知する方法を検討した。ビデオサーバにおける映像データ配信処理を以下に示す。

- (1) ビデオカメラから受信した映像を MPEG2 にて圧縮符号化し、TS パケットを生成する。
- (2) TS パケットの解析を行い、GOP 先頭を含む TS パケットを検出する。
- (3) 解析した TS パケットに GOP 開始コードが存在しない場合は、RTP ペイロードに TS パケットをパッキングする。
- (4) 解析した TS パケットから GOP 開始コードを検出した場合は、パッキング済みの RTP パケットをマルチキャスト配信し、新たに RTP ヘッダを生成する。
- (5) 生成した RTP ペイロードの先頭に GOP 開始コードを含む TS パケットをパッキングし、RTP ヘッダに GOP の先頭を示す識別子を付与する。
- (6) TS パケットが RTP にパッキングできる最大数に達した場合は、生成した RTP パケットをマルチキャスト配信する。

本システムでは、ビデオサーバとゲートウェイとをイーサネットで接続している。イーサネットの

MTUは1500バイト、RTPのヘッダは16バイト、TSパケットは188バイトであることから、1つのRTPにバッキングできるTSパケットの最大数は7個に設定した。IP層におけるフラグメントを避けるためである。

GOPの先頭の有無を示す領域は、既存のRTPヘッダには規定されていないため、本システムにおいて独自に規定する必要がある。拡張ヘッダを定義する方法と、既存の領域使用する方法が考えられる。既存の領域を使用した方が、ゲートウェイにおける処理をより高速化できる。そこで、本システムでは、通常はパディングの有無を示すPビットを、GOPの先頭の有無を示す識別子として用いることとした。

図4にビデオサーバから配信する映像データのパケットフォーマットを示す。①はGOPの先頭を含まないRTPパケット、②はGOPの先頭を含むRTPパケットである。

ビデオサーバで映像の切替点を検出し、RTPのヘッダに識別子をつけて配信するため、ゲートウェイでは切替点の検出処理を高速に行うことができる。

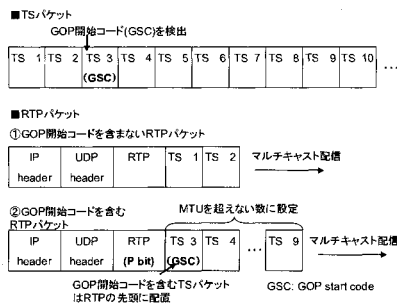


図4 パケットフォーマット

3.2. 切替estream間の同期方法の検討

3.2.1. パーソナライズ映像配信方法

映像連結時の制約を満たしつつ、パーソナライズ映像の送信間隔を一定にする映像配信方法について検討した。切替点がゲートウェイに到達するタイミングのずれを吸収するために、ゲートウェイに各視点の映像を最大1GOP分蓄積するバッファを具備し、切替点でパーソナライズ映像を切替るように設計した。

図5にパーソナライズ映像の切替配信シーケンスを示す。まず、クライアントは、視聴を要求する映像のキーワードをKEY通知メッセージでゲートウェイに登録する。ビデオサーバは、メタデータ通知を用いて配信する映像のキーワードをゲートウ

エイに通知する。ゲートウェイは、時刻 t_1 においてメタデータ通知を受信すると、パーソナライズ映像の切替判定を行う。切替えが必要な場合は、パーソナライズ映像の切替予約を行い、現在配信している映像(GOP1-1)を、切替点を受信するまで続けて配信する。時刻 t_2 においてゲートウェイは、映像の切替点となるGOP2-2の先頭パケットを受信する。しかし、GOP1-1は切替点に達していないため、映像の切替は行わず、GOP1-1の配信を続ける。切替点の判定には、ビデオサーバが設定する識別子(RTPのPビット)を用いる。識別子が付与されたGOP2-2の先頭パケットは、ゲートウェイのバッファに蓄積される。時刻 t_3 において、ゲートウェイは、GOP1-2の先頭のパケットを受信する。本パケットには識別子が付与されているため、映像の切替点と判定する。判定後、直ちにパーソナライズ映像の切替を行う。すなわち、時刻 t_2 よりバッファに蓄積されているGOP2-2の映像の配信を開始する。

図6にゲートウェイにおけるパーソナライズ映像の連結切替を示す。時刻 t_0 から t_3 まではGOP1-1、時刻 t_3 からは時刻 t_2 において受信を開始したGOP2-2がパーソナライズ映像として配信される。

以上により、ゲートウェイは、映像連結点の制約を満たし、パーソナライズ映像の配信間隔を一定に保つことができる。

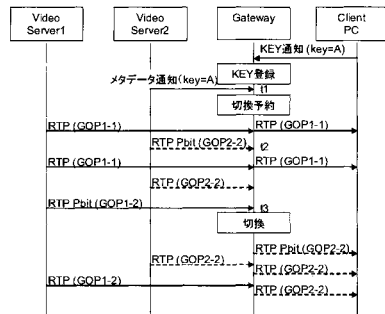


図5 パーソナライズ映像切替シーケンス

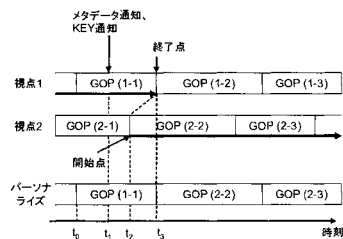


図6 パーソナライズ映像連結切替

3.2.2. タイムスタンプ変換方法

各映像を切換えて1つのストリームに連結する場合、映像のタイムスタンプ情報である、PCR、PTS、DTS(Decode time stamp)が連続するように変換する必要がある。そのため、タイムスタンプ変換方法について検討した。

パーソナライズ映像を GOP 単位で連結した場合、各フレームの PTS、DTS はほぼ一定間隔で増減するため、各視点映像の PTS、DTS に、各視点映像間のオフセットを加えることで求められる。オフセットは、各視点映像の GOP の先頭の PTS の差分から求めることができる。図 7 に PTS/DTS 変換処理を示す。図中には 1 フレームの PTS/DTS のみを示す。パーソナライズ映像切換直後の時刻 t_3 における PTS/DTS_[personalize] の計算式を以下に示す。

$$\text{PTS/DTS}_{[\text{personalize}]} = \text{PTS/DTS}_{[t_1]} + \text{offset}$$

$$\text{offset} = \text{PTS}_{[2]} - \text{PTS}_{[1]}$$

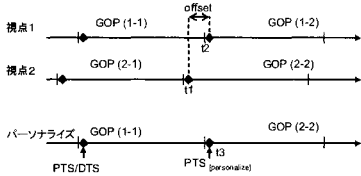


図 7 PTS/DTS 変換処理

一方、PCR の増加間隔は一定ではないため、各視点映像間のオフセットを計算することができない。そこで、映像切換後、最初の PCR を計算するときは、前回 PCR を送信したときの時刻と現時刻との差分から PCR を算出し、その後はオリジナル映像の PCR の差分を前回の PCR の値に加算することで、PCR が連続になるようにした。図 8 に PCR 変換処理を示す。パーソナライズ映像連結直後の時刻 t_3 における PCR_[change] の計算式を以下に示す。

$$\text{PCR}_{[\text{change}]} = \text{PCR}_{[t_1]} + \Delta\text{Time}$$

$$\Delta\text{Time} = t_3 - t_1$$

切換直後以外の PCR_[personalize] の計算式を以下に示す。

$$\text{PCR}_{[\text{personalize}]} = \text{PCR}_{[t_3]} + \Delta\text{PCR}$$

$$\Delta\text{PCR} = \text{PCR}_{[t_4]} - \text{PCR}_{[t_2]}$$

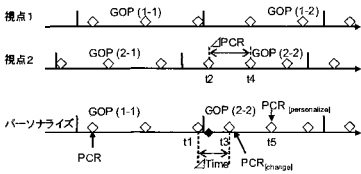


図 8 PCR 変換処理

4. 試作システムの評価

3章で検討した方法に基づいて、本システムの試作を行い、提案方法の有効性を評価した。評価実験における構成機器の仕様を表 1 に示す。

表 1 構成機器仕様

機器	仕様
ビデオサーバ	CPU : Intel Pentium4 3GHz メモリ : 500Mbyte OS : Windows XP
イベントサーバ	CPU : Intel Xeon 1.6GHz メモリ : 2Gbyte OS : Window 2003 Server
ゲートウェイ	CPU : Intel Core2 Duo 2.6GHz メモリ : 3.5Gbyte OS : Fedora 7
クライアント	CPU : Intel Core2 Duo 2.6GHz メモリ : 3.5Gbyte OS : Windows XP SDK : Platform SDK

4.1. クライアントによる映像再生確認

ビデオサーバから配信される各視点の映像と、ゲートウェイで生成されるパーソナライズ映像の再生を目視により確認した。

映像の再生確認には、多視点映像受信クライアントを用いた。多視点映像受信クライアントは、ビデオサーバからマルチキャストで配信される複数の視点映像と、ゲートウェイによって生成されるパーソナライズ映像とを同時に表示することができる。クライアントの表示画面を図 9 に示す。

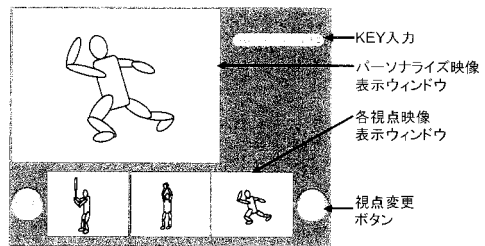


図 9 クライアント表示画面

ゲートウェイから配信されるパーソナライズ映像は、画面左上のパーソナライズ映像表示ウィンドウに表示される。ビデオサーバから配信される複数の視点映像は、画面下部の視点映像表示ウィンドウに表示される。ウィンドウに表示される映像は、左右の視点変更ボタンで変更することができ、3視点以上の映像を受信する場合は、表示、非表示を順番に切替えることができる。右上の KEY 入力ウィン

ドウから、メタデータとして登録されるキーワードのひとつを入力することができ、ユーザが要求する映像のキーワードをゲートウェイに通知することができる。

ビデオサーバから5本の視点映像を配信し、各視点の映像と、ゲートウェイで生成されるパーソナライズ映像の表示を確認した。多視点映像受信クライアントでは、各視点映像とパーソナライズ映像を同時に生成可能であることを確認した。また、各視点映像を切替えて生成したパーソナライズ映像をノイズなく滑らかに再生可能であることを確認した。

4.2. 最大収容ユーザ数の評価

1台のゲートウェイで収容可能なユーザ数を評価するために、ゲートウェイが配信するストリームのバケット転送レートを計測した。擬似クライアントからパーソナライズ映像配信要求を送信し、ゲートウェイから複数のパーソナライズ映像を配信する。ユーザ数を5ずつ増加させ、ゲートウェイにおける1ストリームあたりのバケット転送レートを計測した。映像品質については目視により確認した。1ストリームあたりのバケット転送レートを図10に示す。各視点の映像は、1ストリーム平均464pps(packet per second)でビデオサーバから配信されている。

RTPヘッダに付与された識別子を用いずに切替点を検出する方式では、ユーザ数40程度で大幅な転送レートの低下を確認した。また、目視による評価では、映像にブロックノイズや再生が断続的になる等の品質低下が見られた。一方、切替点の検出に識別子を用いた方式では、大幅な転送レートの低下は確認されなかった。また、目視による評価では、ユーザ数90以下の場合にはブロックノイズ等の映像品質の低下は見られなかった。

識別子を利用した方式は、識別子を利用しない方式に比べて、約3倍のユーザを収容可能であることを確認した。

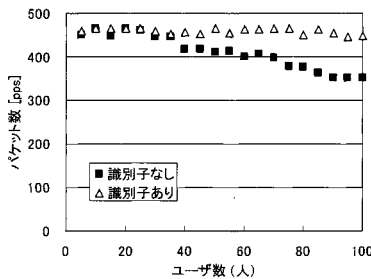


図10 パケット転送レート

4.3. ゲートウェイにおけるCPU使用率の評価

ビデオサーバから5本の視点映像を配信し、5本または40本のパーソナライズ映像を生成した時のゲートウェイにおけるCPU使用率を計測した。計測結果を図10に示す。

RTPヘッダに付与された識別子を用いずに切替点を検出する方式のCPU使用率は、それぞれ6%、22%であった。一方、識別子を用いた方式のCPU使用率は、それぞれ1%、7%であった。

識別子を利用した方式は、識別子を利用しない方式の1/3以下のCPU使用率でパーソナライズ映像が生成可能であることを確認した。

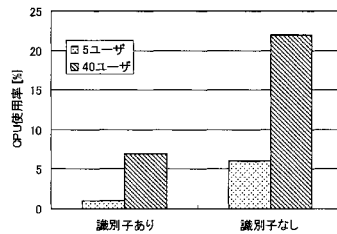


図11 ゲートウェイにおけるCPU使用率

5. 結論

本研究では、ビデオサーバから配信される複数の視点の映像の中から、視聴者が要求するキーワードに対応した映像を選択し、ゲートウェイにて映像を切替えて配信するパーソナライズ映像配信システムを提案した。本システムのゲートウェイにおける課題について検討し、映像切替点の検出処理を低減する方法及び、非同期に配信される映像ストリームから切替点の制約を満たしたパーソナライズ映像を生成する方法について検討した。検討した方法に基づいて、パーソナライズ映像配信システムの試作を行い、ゲートウェイにおけるパーソナライズ映像配信性能の評価を行った。その結果、ゲートウェイで生成したパーソナライズ映像を、クライアントではノイズなく滑らかに再生可能であることを確認した。また、1台のゲートウェイで90ユーザにパーソナライズ映像を配信可能であることを確認した。

文献

- [1] <http://www.doing.tv/>
- [2] H. Ikeda, J. Sugawa, Y. Ashi and K. Sakamoto, "High-definition IPTV Broadcasting Architecture over Gigabit-capable Passive Optical Network," Globecom, November 2007
- [3] 谷本正幸, "自由視点テレビFTV", 電子情報通信学会論文誌 A, vol.89, no.11, pp.866-872, Nov. 2006