

複数カメラからの映像の伝送順序が 映像品質に与える影響について

小寺 志保¹ 藤橋 卓也^{2,3} 猿渡 俊介¹ 渡辺 尚²

概要: 本稿では、無線通信上で低トラフィック、高品質なマルチビュービデオ伝送を達成する伝送方式として Domino Streaming++ を提案する。Domino Streaming++ では、複数の撮影機器が互いに無線通信を傍受し、得られた他の撮影機器の映像をエンコードに利用することで、トラフィック削減を達成する。また、撮影機器間の映像の類似度に基づいて、各撮影機器の送信順を制御することで、より大幅なトラフィック削減を達成する。MERL が提供している実ビデオシーケンスを用いた評価から、Domino Streaming++ のトラフィック量は無線マルチビュービデオの伝送におけるトラフィック量の理想値に漸近することが分かった。

1. はじめに

立体映像、自由視点映像 [1, 2] に代表される超臨場感映像の発展は、話し手と聞き手の間にある物理的距離を限りなくゼロに近づける。例えば、テレプレゼンス [3, 4] が実現されることで遠隔にいる人々でも face-to-face と同等の円滑なコミュニケーションを提供できる。オリンピック種目の自由視点映像配信では、会場さながらの熱狂を遠隔地で味わうことができるエンターテインメントを提供できる。高度道路交通システムでは、複数の車載カメラを用いて高度な安全運転支援を実現できる。

マルチビュービデオは、超臨場感映像の発展を担う主要技術の 1 つであり、ある被写体を近隣に設置した複数台のカメラで同時に撮影し、撮影した複数の映像を視聴者に提供する技術である。

図 1 にマルチビュービデオを撮影してから、視聴者が映像を視聴するまでの構成を示す。マルチビュービデオシステムは、撮影部、転送部、表示部の 3 種類から構成される。

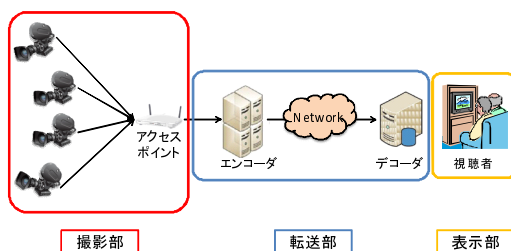


図 1 マルチビュービデオシステムの構成

撮影部では、被写体を複数台のカメラを用いて同時に撮影し、エンコーダまで撮影した映像を伝送する。転送部では、撮影部から受信した映像をエンコードし、エンコードした映像を視聴者のデコーダまで伝送する。表示部では、転送部から届いた映像をデコードし、表示機器上に映像を表示する。

転送部の代表的な研究として、H.264/AVC によって標準化された Multi-view Video Coding (MVC) [5-7]、YouTube など蓄積型配信に対応した Interactive Multi-view Video Streaming (IMVS) [8-12]、ライブ配信に対応した User Dependent Multi-view Video Transmission (UDMVT) [13-15]、複数の視聴者への配信に対応した User dependent Multi-view video Streaming for Multi-user (UMSM) [16-18] が挙げられる。表示部の代表的な技術には、インテグラル立体テレビ [19]、自由視点テレビ [2] がある。

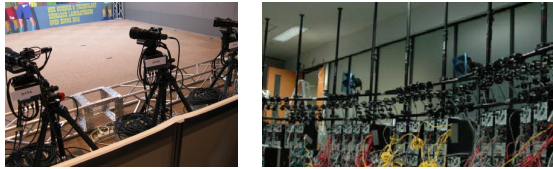
一方で、現在のマルチビュービデオの撮影部では図 2 に示すマルチカメラアレーが用いられている。図 2(a) は直線上にカメラを配置したカメラアレー [21] であり、図 2(b) は円形にカメラを配置したカメラアレー [20] である。しかしながら、マルチカメラアレーは有線ネットワークで繋がれた複数台のカメラを利用することを想定しているため、カメラを設置できる環境が限定的になり、マルチビュービデオを展開する際の障壁となっている。

本研究では、デジタルカメラやスマートフォンといった小型撮影機器と無線通信を組み合わせることで、場所を選ばずマルチビュービデオが撮影できるシステムを考える。撮影部を無線化するためには、2 つの要件を満たす必要がある。1 つ目の要件は、トラフィックの削減である。

¹ 静岡大学 大学院 情報学研究科

² 大阪大学 情報科学研究科

³ 日本学術振興会特別研究員 (DC1)



(a) ロボットカメラ (b) 自由視点映像の撮影システム [20]

図 2 既存マルチビュービデオ撮影部

マルチビュービデオが複数台のカメラの映像から構成されるのに対して、無線通信は電波という有限の資源を共有するため有線通信よりデータレートが低い。2つ目の要件は、映像品質の維持である。これら2つの要件を同時に満たさなければ、アプリケーションに対する視聴者満足度が低下する。

本稿では、撮影部を無線化した際に、トラヒックの削減と映像品質維持の2つの要件を満たす方式として、Domino Streaming++を提案する。Domino Streaming++では、ある撮影機器は他の撮影機器が送信しているビデオを傍受し、傍受した映像と自身の映像との差分をエンコードすることでトラヒックを削減する。また、撮影機器間にある冗長な情報をより多く削減するため、アクセスポイントが撮影機器間の映像の類似度に従って送信順を制御する。JMVCエンコーダ上に実装した計算機シミュレーションとMitsubishi Electronics Research Laboratory (MERL) が提供する実ビデオシーケンスを利用してDomino Streaming++の性能を評価した。評価結果から、他の撮影機器の映像を利用しない場合と比較して、同じ映像品質下において、Domino Streaming++が最大46%のトラヒック量を削減できることが分かった。同時に、Domino Streaming++が無線マルチビュービデオ伝送におけるトラヒック量の理想値に漸近する性能を達成することが分かった。

本稿の構成は以下のとおりである。2節では、撮影部の無線化を実現するための要件について述べる。3節では、本研究の関連研究について述べる。4節では、提案方式であるDomino Streaming++について述べる。5節では、JMVCエンコーダと実ビデオシーケンスによって、提案方式であるDomino Streaming++の性能を評価する。最後に6節でまとめとする。

2. 本研究の位置づけ

本研究では、マルチビュービデオをより多くの場面で利用できるようにするため、撮影部を無線化したマルチビュービデオシステムの実現を目指す。撮影部を無線化することで、スタジオなどの屋内に限られた場所だけでなく、より多くの被写体をマルチビュービデオで提供することが可能となる。例えば、屋外イベントの際にイベントの状況

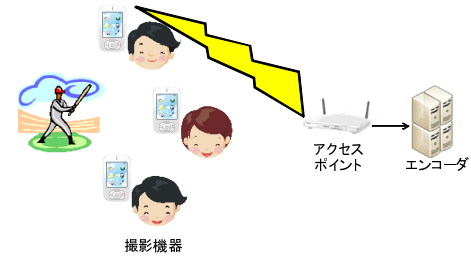


図 3 撮影部の無線化モデル

を各参加者が持っているスマートフォン等の小型撮影機器で撮影し、その場にはいないユーザに対してマルチビュービデオとして提供することができる。

図3に、無線化した撮影部の通信モデルを示す。アクセスポイント間は無線によって接続され、アクセスポイントからエンコーダまでは有線によって接続されている。各撮影機器は自身が撮影した映像をアクセスポイントまで伝送し、アクセスポイントは複数の撮影機器から受信した映像をエンコーダへ伝送する。無線化したマルチビュービデオの撮影部を想定した場合、視聴者が、実際の被写体に限りなく近い映像を不自然に途切れることなく視聴できることが必要となる。そのためには、次の2つの要件を同時に満たす必要がある。

1つ目の要件は、トラヒックを少なくすることで、伝送遅延による視聴者満足度の低下を防ぐことである。マルチビュービデオは複数の撮影機器からの映像を扱うため、シングルビデオよりトラヒックが多くなる。一方で、無線通信では、有限である電波を複数のノードによって共有しているため、有線通信と比べると伝送レートが低い。単純にマルチビュービデオを無線通信で伝送した場合、伝送レートの低さによって映像を撮影してから視聴者へ映像を届けるまでの遅延が大きくなり、結果として視聴者満足度が低下する。例えば、視聴者が視点を切り替えたときに、所望の映像が届くことが遅延していることに起因して、映像が停止したりスキップしたりする。

2つ目の要件は、映像品質を高く維持することである。映像品質とは、元の映像とデコードした映像との劣化の度合いを表す。品質が高い映像とは、劣化が小さく、再現率の高い映像を意味する。視聴者は、実際との被写体と同等の映像を求めるため、映像品質の低下は視聴者満足度の低下を招く。

1つ目の要件であるトラヒックの削減と2つ目の要件である映像品質の維持はトレードオフの関係にある。最も単純に撮影部の無線化を実現する方法として、各撮影機器が無線通信で映像を個別に送信することが考えられる。しかしながら、各撮影機器が送信する複数の映像によるトラヒックが1つの無線通信の伝送レートを共有することになるため、1つの撮影機器あたりの伝送レートが低くなる。その結果、視聴者への映像の伝送に遅延が発生し、視聴者

満足度が低下する。

トラフィックを削減する単純な方法として、マルチビュービデオの量子化パラメータを大きくすることやフレームレートを落とすことが考えられる。量子化パラメータとは、映像を量子化する際に元映像の情報をどれだけ残すかを表すパラメータである。例えば、マルチビュービデオをエンコードするとき量子化パラメータを大きくした場合、量子化によって多くの映像データが捨てられるため、トラフィック量が減少する。一方で、映像をデコードするとき、捨てられた映像データの欠損によって元の映像データを復元しにくくなるため、映像品質が劣化する。

3. 関連研究

本研究は映像間の類似度算出技術と分散マルチビュービデオ符号化技術に関係する。

類似度算出技術には、ビデオフレームの画素値から得られる特徴量を元に算出する方式とビデオフレーム内の特徴点を元に算出する方式がある。ビデオフレームの画素値を用いる方式として Information-Bound References (IBR) [22, 23] が挙げられる。IBR を用いることで、ビデオフレームの画素値を元に、92 [byte] で表される映像の特徴量を算出できる。その後、2 映像の IBR 値を比較することで、2 映像の類似度を算出できる。

ビデオフレーム内の特徴点を用いる方式として Scale-Invariant Feature Transform (SIFT) [24] が挙げられる。SIFT はあるビデオフレームに対して、回転、スケールの変化、明度の変化などの操作を加えても変化が現れない特徴点を算出する。その後、2 枚のビデオフレームと各ビデオフレーム内の特徴点を用いることで、映像間の類似度を算出できる。Domino Streaming++ では、各撮影機器がアクセスポイントに対して IBR を元にした特徴量を送信した場合、撮影機器間の映像類似度に基づいたトラフィック削減を達成する。IBR を元にした特徴量を利用した場合、フレーム内の特徴点を用いる場合と比較して、各撮影機器は類似度算出のために自身のビデオフレームを送信する必要はなく、より少ない伝送遅延でアクセスポイントに対して自身の特徴量を伝送できる。

マルチビュービデオにおいて、映像品質を維持したまま無線通信時のトラフィックを削減する方法として、Distributed Multiview Video Coding (DMVC) [25–27] が挙げられる。DMVC は、マルチビュービデオの無線伝送において、映像間の相関を利用して符号化レベルでアプローチしている研究である。DMVC では、分散情報源符号化を適用すると共に、撮影機器の設置位置や撮影しているアングルを映像データと共に side information として送信する。分散情報源符号化とは、相関を有する 2 つのデータに独立に符号化しても参照した時と同程度の圧縮率を達成できる符号化手法であり、Slepian-Wolf 理論 [28] や Wyner-Ziv 理論 [29]

表 1 前提条件

層	プロトコル
トランスポート	UDP
ネットワーク	IP
MAC・物理	IEEE 802.11a

などが存在する。Domino Streaming++ は映像の前処理部分に着目したアプローチであるため、符号化レベルでのアプローチである DMVC と組み合わせて使用することもできる。

4. Domino Streaming++

無線化したマルチビュービデオ撮影部において低トラフィックと映像品質の維持を達成する Domino Streaming++ を設計した。Domino Streaming++ では、無線通信において近隣のノードが他のノードの通信を受信できるという特性を利用して、傍受した映像の差分を取ることで高い圧縮率を実現する。

表 1 に Domino Streaming++ が各層で用いることを想定しているプロトコルを示す。物理層・MAC 層は IEEE 802.11a [30]、ネットワーク層は IP、トランスポート層は UDP を用いることを想定している。

4.1 動作の全体像

Domino Streaming++ は初期化、特徴量算出、送信順決定、エンコード、映像伝送、デコードの流れで動作する。

- (1) Domino Streaming++ に参加する撮影機器は、アクセスポイントの通信範囲内に入ると初期化を開始し、アクセスポイントによって送信順が割り当てられる。初期化の詳細については 4.2 節で述べる。
- (2) 各撮影機器は初期化を開始すると、自身の映像の特徴量を、ビデオフレームの画素値を元に算出する。特徴量を算出すると、各撮影機器は算出した特徴量をアクセスポイントに対して送信する。特徴量算出の詳細については 4.3 節で述べる。
- (3) 各撮影機器の初期化が終了すると、アクセスポイントは初期化中に取得した各撮影機器の映像の特徴量から映像送信順を決定する。映像送信順を決定すると、アクセスポイントは各撮影機器に決定した送信順をブロードキャストする。送信順決定の詳細については 4.4 節で述べる。
- (4) 各撮影機器はアクセスポイントが決定した送信順に従って、自身の映像をすでに傍受した他の撮影機器の映像を用いて一定量エンコードする。エンコードの詳細については 4.5 節で述べる。
- (5) 各撮影機器はアクセスポイントが決定した送信順に従ってエンコードした自身の映像をアクセスポイントに対して伝送する。また、他の撮影機器は通信を傍

受することで、送信中の撮影機器の映像を受信する。全ての撮影機器が映像を送信し終わると、アクセスポイントは再び各撮影機器の映像の特微量に基づいて、映像送信順序を決定する。映像伝送の詳細については4.6節で述べる。

- (6) 各撮影機器とアクセスポイントから映像を受信しているエンコーダは、新たに撮影機器の映像を受信するとデコードを開始する。デコードの詳細については4.7節で述べる。

4.2 初期化

Domino Streaming++では、各撮影機器は映像を伝送する前に、アクセスポイントによって一意のIDを割り当てられる。アクセスポイントはアクセスポイントの通信範囲に入ったことを撮影機器に通知するビーコンを周期的に送信する。撮影機器はビーコンを受信すると、IDを取得するために、自身が撮影している映像の特微量をアクセスポイントに対して送信する。映像の特微量の算出方法については、4.3節で述べる。アクセスポイントは各撮影機器の映像の特微量を受信すると、フレームを送信した撮影機器に対してID割り当てフレームを送信する。ID割り当てフレームにはアクセスポイントが各撮影機器に割り当てた一意のIDが格納される。

4.3 特微量算出

各撮影機器はアクセスポイントの通信範囲内に入ると、自身の映像の特微量を算出し、アクセスポイントに対して送信する。映像の特微量はInformation-Bound References (IBR) [22, 23]に基づいて算出する。映像の特微量は、映像を構成する各ビデオフレームの周波数成分を利用して算出する。周波数成分は、各ビデオフレームに含まれる画素値をDCT変換することで求められ、大きく高周波成分と低周波成分の2種類に分けられる。高周波成分はビデオフレームの詳細な特徴を表している。高周波成分が変化しても、人間はほとんどビデオフレームに対する変化を視認できない。一方で、低周波成分はビデオフレームの大まかな特徴を表している。低周波成分が変化すると、ビデオフレームには人間が視認できる変化が生じる。

各撮影機器は、自身が撮影した映像の1 Group of Pictures (GOP) 分のビデオフレームのうち、先頭のビデオフレームから特微量を算出する。GOPとは、複数のビデオフレームの集合体であり、通常は8枚のビデオフレームから構成される。まず、各撮影機器は先頭のビデオフレームの解像度を128×128画素に変換する。その後、撮影機器はビデオフレームの各画素値を4:2:2の輝度信号Y、色差信号Cb、Crに変換する。ビデオフレームの各画素値は合計24ビットのRGB信号から構成されているが、4:2:2の輝度信号及び色差信号に変換することで、RGB信号と

表 2 Algorithm 1 で使用する変数、関数

変数、関数	説明
C	アクセスポイントの通信範囲内にある撮影機器のIDの集合
$order[i]$	i 番目に映像を送信する撮影機器のIDを格納する配列
$\arg \min_{a \in C} f(a)$	C に属する撮影機器のうち $f(a)$ が最小となる撮影機器 a のIDを取得する関数
I_b	撮影機器 b の映像の特微量
$calc_hamingdistance(I_c, I_d)$	特微量 c と d の間のハミング距離を算出する関数
$delete(C, e)$	撮影機器の集合 C からIDが e である撮影機器を削除する関数
$size(C)$	C に含まれる元の個数

同じ色分解能を保ったまま、各画素値を8ビットの輝度信号Y、4ビットの色差信号Cr、4ビットの色差信号Cbの合計16ビットで表現することができる。撮影機器は輝度信号Yに対してDCT変換を実行し、輝度信号の低周波成分と高周波成分を取得する。

各撮影機器は輝度信号Yの低周波成分から、自身の映像の特微量を求める。具体的には、128×128画素の輝度信号Yのうち、先頭の8×8画素の低周波成分を取り出す。撮影機器は取り出した合計64個の輝度信号Yの低周波成分から中央値を算出する。その後、取り出した各周波数成分が中央値を超えているかに応じて、1または0を割り当てる。最終的に、各撮影機器は特微量として64個の0、1列を取得する。

4.4 送信順決定

アクセスポイントは各撮影機器から受信した映像の特微量を元に、各撮影機器の映像送信順を決定する。具体的には、各撮影機器が自身の映像と最も類似した映像を利用して、映像をエンコード可能な映像送信順を決定する。アクセスポイントは撮影機器間の映像類似度を求めるために、2台の撮影機器の特微量からハミング距離を算出する。ハミング距離が小さいほど、2台の撮影機器の映像が類似することを示している。

Algorithm 1 に撮影機器が N 台の場合の映像類似度に基づく送信順決定アルゴリズムを示す。送信順決定アルゴリズムは送信開始機器の決定と後続機器の送信順決定の2つから構成される。表2に、Algorithm 1 で使用する変数と関数を示す。 C はアクセスポイントの通信範囲内にある撮影機器のIDの集合である。 $order[i]$ は i 番目に映像を送信する撮影機器のIDを格納する配列である。 $\arg \min_{a \in C} f(a)$ は C に属する撮影機器のうち $f(a)$ が最小となる撮影機器 a のIDを取得する関数である。 I_b は撮影機器 b が撮影している映像の特微量を表す。 $calc_hamingdistance(I_c, I_d)$

Algorithm 1 Transmission order control

```
1:  $i \leftarrow 1$ 
2:  $x \leftarrow 1$ 
3:  $order[i] \leftarrow x$ 
4:  $i \leftarrow i + 1$ 
5: delete ( $C, x$ )
6: while size ( $C$ ) > 0 do
7:    $y \leftarrow \arg \min_{y \in C} (\text{calc\_hammingdistance} (I_y, I_x))$ 
8:    $order[i] \leftarrow y$ 
9:    $i \leftarrow i + 1$ 
10:   $x \leftarrow y$ 
11:  delete ( $C, y$ )
12: end while
```

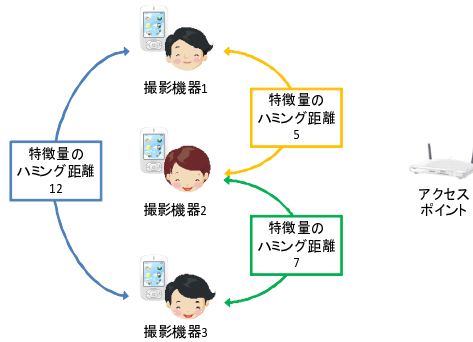


図 4 撮影機器が 3 台のときの配置例

は特徴量 c と特徴量 d の間のハミング距離を算出する関数である。delete (C, e) は撮影機器の集合 C から ID が e である撮影機器を削除する関数である。size (C) は C に含まれる元の個数であり、初期値は撮影機器数 N である。

1 つ目の処理である開始機器の決定では、 N 台の撮影機器の中で、最初に映像を送信する撮影機器を決定する。Domino Streaming++ では、アクセスポイントに ID 1 を割り当てられた撮影機器を開始機器とする。具体的には、Algorithm 1 の 1 行目から 5 行目で、アクセスポイントの通信範囲内にある撮影機器のうち、ID 1 が割り当てられた撮影機器 x を開始機器とする。

2 つ目の処理である後続機器の送信順決定では、開始機器を除く全ての撮影機器の送信順を決定する。具体的には、撮影機器間の特徴量のハミング距離が最小になる映像送信順を決定する。Algorithm 1 の 7 から 11 行目で、まだ送信順を割り当てていない撮影機器の映像の特徴量の中で、直前に送信順を割り当てた撮影機器 x の映像の特徴量とのハミング距離が最小となる撮影機器 y を選択する。アクセスポイントは撮影機器 y に送信順を割り当て、撮影機器 y を撮影機器 x とみなし、Algorithm 1 の 7 から 11 行目の処理を繰り返す。

例として 3 台の撮影機器がある場合を考える。この時、アクセスポイントは撮影機器 1 に ID 1、撮影機器 2 に ID 2、撮影機器 3 に ID 3 を割り当て、撮影機器間の特徴量のハミング距離が図 4 のようになったと仮定する。

表 3 ビデオパケット $P_{i,j}$

フィールド	バイト
映像類似度	8
映像	可変長

まず、アクセスポイントは ID 1 が割り当てられた撮影機器 1 を開始機器とし、order[1] に撮影機器 1 の ID を格納する。次に、撮影機器 1 の映像の特徴量と撮影機器 2、3 の映像の特徴量とのハミング距離を算出する。撮影機器 1 と撮影機器 2 の間のハミング距離は 5、撮影機器 1 と撮影機器 3 の間のハミング距離は 12 であるため、アクセスポイントはハミング距離がより小さい撮影機器 2 の ID を order[2] に格納する。その後、アクセスポイントは残った撮影機器 3 の ID を order[3] に格納し、送信順決定アルゴリズムを終了する。最終的に、アクセスポイントが決定した送信順は撮影機器 1 → 撮影機器 2 → 撮影機器 3 となる。

4.5 エンコード

送信順が決定すると、各撮影機器は送信順に従って自身の映像をエンコードする。各撮影機器は H.264/AVC に基づき、GOP ごとに自身の映像をエンコードする。Domino Streaming++ では、撮影機器は自身の映像を低トラフィックで送信するために、自身に割り当てられた送信順より前の通信を全て傍受する。

図 5 から図 7 に GOP のエンコードを示す。ここで撮影機器の台数は 3 台であり、映像送信順は撮影機器 1 → 撮影機器 2 → 撮影機器 3 の順であると想定する。図 5 に、撮影機器 1 の予測構造を示す。撮影機器 1 は単独でエンコードするため、先頭のフレームが I フレームとなる。I フレームは、JPEG によって符号化された静止画情報であり、データ量が大きい。

図 6 に、撮影機器 2 の予測構造を示す。撮影機器 2 は、撮影機器 1 の映像を傍受しているため、撮影機器 1 と自身の映像との差分情報をエンコードする。具体的には、撮影機器 2 の先頭フレームを P フレームとしてエンコードする。P フレームは他の撮影機器の I フレームを元に作成した自身の映像との差分情報であるため、I フレームよりデータ量が少なくなる。

図 7 に撮影機器 3 の予測構造を示す。撮影機器 3 は、撮影機器 1 と 2 の映像を傍受しているため、撮影機器 1、2 の映像と自身の映像との差分情報をエンコードする。具体的には、撮影機器 3 の先頭フレームを P フレームとしてエンコードする。

4.6 映像伝送

各撮影機器はアクセスポイントが割り当てた送信順に従って、4.5 節に示した方法でエンコードした映像を伝送する。各撮影機器は 1 GOP ずつ自身の映像をアクセスポ

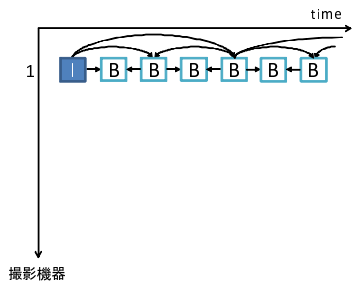


図 5 撮影機器 1 の映像のエンコード

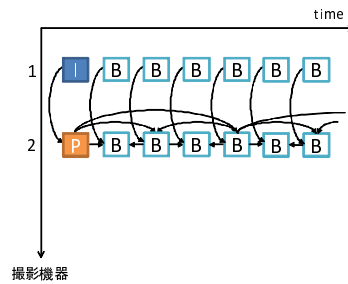


図 6 撮影機器 2 の映像のエンコード

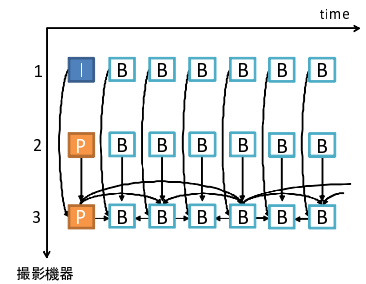


図 7 撮影機器 3 の映像のエンコード

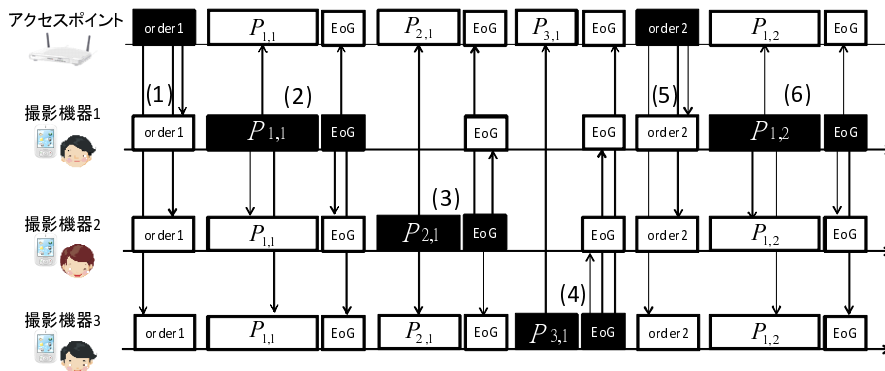


図 8 映像伝送タイムシーケンスチャート

イントに対して送信する．図 8 に映像伝送のタイムシーケンスチャートを示す．図 8 では，アクセスポイントが撮影機器 1→撮影機器 2→撮影機器 3 の順に送信順を決定したと仮定する．また， $P_{i,j}$ は，撮影機器 i の GOP j のビデオフレームから構成される packets を表している．

表 3 に $P_{i,j}$ のフォーマットを示す．映像類似度フィールドには，4.4 節で述べた方法に従って算出した各撮影機器の映像の特徴量を格納する．映像フィールドには，撮影機器 i がエンコードした GOP j の映像を格納する．

- (1) アクセスポイントは 4.4 節の方法に従って決定した GOP1 に対する各撮影機器の送信順を全ての撮影機器にブロードキャストする．
- (2) アクセスポイントから受信した送信順に基づいて，撮影機器 1 から映像の伝送を開始する．撮影機器 1 はエンコードした GOP1 の映像を映像フィールド，GOP2 の先頭のビデオフレームから算出した特徴量を映像類似度フィールドに格納し， $P_{1,1}$ をアクセスポイントに対して送信する．撮影機器 2 と 3 は $P_{1,1}$ を傍受して撮影機器 1 の映像のデコードを開始する．撮影機器 1 は $P_{1,1}$ を送信し終わると，続けて EoG (End of GOP) パケットを送信する．EoG パケットは，1 GOP 分の映像の送信が終了したことを他の撮影機器に通知するために利用され，IEEE 802.11 における ACK フレーム [30] と同じフォーマットを利用する．EoG パケットを傍受した撮影機器 2 は，次が自身の送信順だと判断し，デコードが終了した撮影機器 1 の映像を利用して自身の映像のエンコードを開始する．一方，撮影機

器 3 は，次はまだ自身の送信順ではないと判断し，デコードが終了した撮影機器 1 の映像を保持し，撮影機器 3 の通信が終了するまで待機する．

- (3) 撮影機器 2 は，エンコードした GOP1 の映像を映像フィールド，GOP2 の先頭のビデオフレームの特徴量を映像類似度フィールドに格納し，アクセスポイントに対して $P_{2,1}$ を送信する．撮影機器 3 は $P_{2,1}$ を傍受することで撮影機器 2 の映像を取得する．その後，撮影機器 3 は受信した映像のデコードを開始する．撮影機器 2 は $P_{2,1}$ の送信が終了すると，EoG パケットをアクセスポイントへ送信する．EoG パケットを傍受した撮影機器 3 は，次が自身の送信順だと判断し，デコードが終了した撮影機器 2 の映像と保持していた撮影機器 1 の映像を利用して自身の映像をエンコードする．
- (4) 撮影機器 3 は，エンコードした GOP1 の映像を映像フィールド，GOP2 の先頭のビデオフレームの特徴量を映像類似度フィールドに格納し，アクセスポイントに対して $P_{3,1}$ を伝送する．撮影機器 3 は $P_{3,1}$ の送信が終了すると，EoG パケットをアクセスポイントへ送信する．
- (5) アクセスポイントは受信した $P_{1,1}$ ， $P_{2,1}$ ， $P_{3,1}$ を有線通信によってそのままエンコーダへ転送すると同時に， $P_{1,1}$ ， $P_{2,1}$ ， $P_{3,1}$ に含まれている各撮影機器の特徴量から，4.4 節の方法に従って，GOP2 に対する映像伝送順を決定する．その後，決定した送信順を全ての撮影機器にブロードキャストする．
- (6) GOP2 に対する映像送信順を受信した各撮影機器は，

映像送信順に従って GOP2 に対する映像伝送を開始する。ここでは、GOP2 に対する映像送信順が撮影機器 1→撮影機器 2→撮影機器 3 の順であると仮定する。GOP1 のときと同様に、撮影機器 1 はエンコードした GOP2 の映像を映像フィールド、GOP3 の先頭のビデオフレームの特徴量を映像類似度フィールドに格納し、 $P_{1,2}$ をアクセスポイントに対して送信する。

Domino Streaming++ではすべての GOP の送信が終了するまで、(2) から (6) の動作を繰り返す。

4.7 デコード

Domino Streaming++のデコードには特別な処理を必要としない。撮影機器とアクセスポイントから各撮影機器の映像を受信しているエンコーダは標準的な H.264/AVC デコーダを利用して、それぞれ受信した映像をデコードする。各撮影機器とエンコーダが最初に受信する映像は、先頭フレームが静止画である I フレームによってエンコードされているため、撮影機器とエンコーダは映像を受信すると同時にデコードを開始する。1 番目に送信された撮影機器の映像以降、撮影機器とエンコーダが受信する映像は、すでに受信した映像を元に作成された差分映像である。撮影機器とエンコーダはすでに受信した映像のデコードが終了すると同時に、新たに受信した映像のデコードを開始する。エンコーダでは全ての撮影機器の映像をデコードすると、転送部で利用する方式に従って全ての撮影機器の映像を再びエンコードする。最後に、エンコードした映像を視聴者に対して伝送することで、視聴者は複数台の撮影機器が撮影したマルチビュービデオを視聴することができる。

5. 性能評価

5.1 評価環境

Domino Streaming++の有効性を確認するために、JMVC エンコーダ [31] 上に実装した計算機シミュレーションと MERL が提供しているテストビデオシーケンス [32] によって、各撮影機器の映像に対する特徴量と映像品質に対する映像ビットレートを測定した。Domino Streaming++の性能を相対的に評価するために、次の 3 つのアプローチを比較した。

(1) Independent Streaming

Independent Streaming は、各撮影機器が無線通信で自身の映像をアクセスポイントに対して個別に送信する方式である。Independent Streaming は、他の撮影機器の通信を傍受しないため、Domino Streaming++において他の撮影機器の映像を利用することによって得られた性能を示す尺度となる。

(2) Ideal

Ideal は各撮影機器が理想的な組み合わせでエンコードした場合のトラフィック量である。Ideal は無線マル

表 4 評価パラメータ

解像度	176×144
フレームレート	15 fps
フレーム数	250
GOP サイズ	8 フレーム
カメラ数	8 台
量子化パラメータ (QP)	24~40

チビュービデオにおけるトラフィック量の理想値である。Ideal では、各撮影機器が撮影している映像は全知であると仮定する。アクセスポイントは各撮影機器の映像に応じて、GOP ごとに理想的な映像送信順を決定する。Ideal は、Domino Streaming++が無線マルチビュービデオ伝送における性能限界にどの程度近づいたかを示す尺度となる。

(3) Domino Streaming++

Domino Streaming++は、4 節で述べた提案方式である。Domino Streaming++では、近隣の撮影機器の映像を利用して自身の映像をエンコードすることで、トラフィックを削減する。また、アクセスポイントによる送信順制御によって、より大幅なトラフィック削減を達成する。

共通の評価パラメータを以下に示す。テストビデオシーケンスには、Ballroom, Exit, Vassar を利用した。Ballroom は舞踏会の映像で、動きが多い。Exit は出入り口の映像で、Ballroom ほどではないが動きがある。Vassar は風景の中を車が走る映像で、動きは少ない。撮影機器とアクセスポイント間の無線通信路はロスレスを想定した。Ballroom, Exit, Vassar における各撮影機器間の距離は 19.5 cm である。

表 4 に JMVC エンコーダのパラメータを示す。フレームレートは 15 fps、各テストビデオシーケンスのフレーム数は 250 フレームである。1 GOP あたりのフレーム数は 8 フレームとし、撮影機器数は 8 台とした。各評価にあたって、量子化パラメータを 24 から 40 まで変化させ、異なる Peak Signal-to-Noise Ratio (PSNR) 時のトラフィック量を取得した。PSNR とは映像品質を表すために一般的に用いられている指標であり、次式で表される。

$$\text{PSNR} = 20 \log_{10} \left(\frac{\text{MAX}}{\sqrt{\text{MSE}}} \right)$$

MAX は元画像がとりうる最大画素値、MSE は元映像とデコード後の映像との平均二乗誤差 (Mean Square Error) である。

5.2 映像の特徴量算出

マルチビュービデオにおける映像類似度の特性を評価するために、4.3 節に示した方法で撮影機器間の映像類似度を評価した。

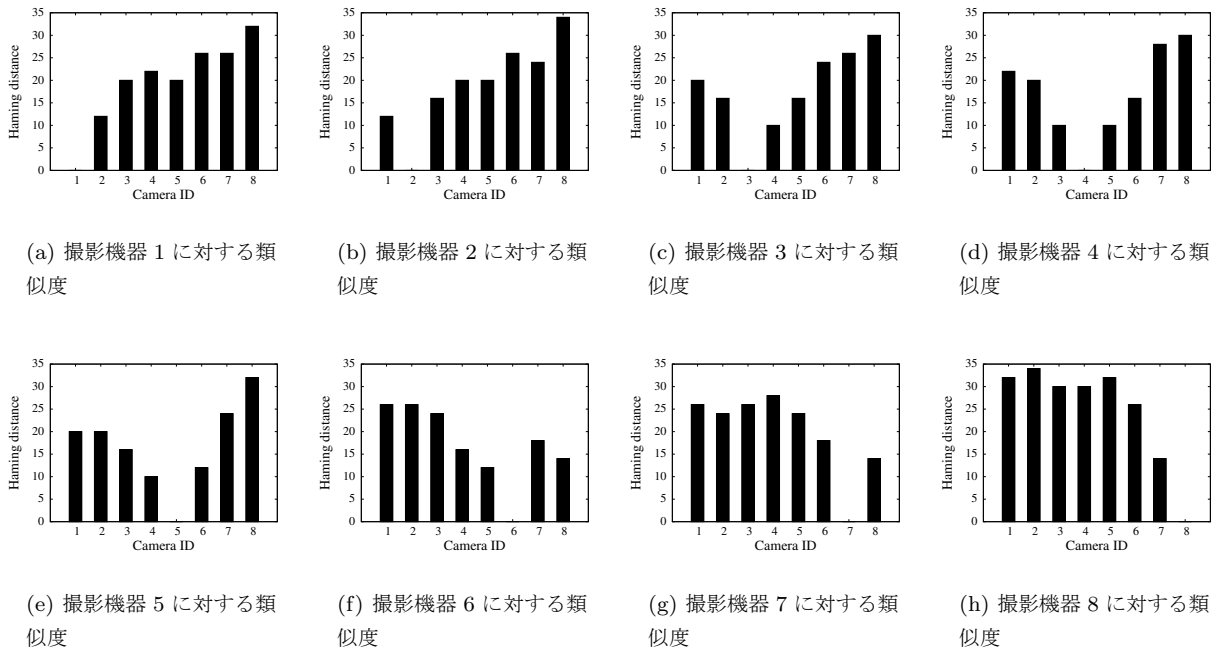


図 9 特徴量のハミング距離

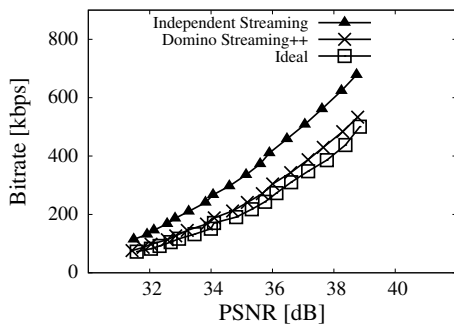


図 10 理想的な映像ビットレートとの比較

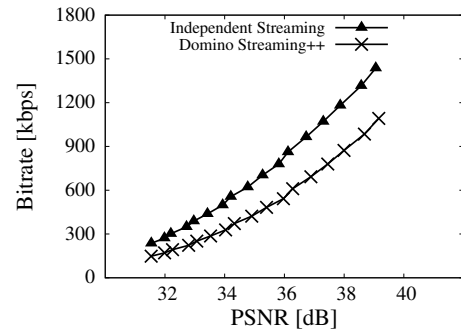


図 11 映像品質に対する映像ビットレート (Ballroom)

図 9 に、利用したビデオシーケンスが Ballroom、各撮影機器の映像の中で GOP1 の先頭フレームを特徴量算出に使用した場合の 2 撮影機器間の映像類似度を示す。図 9(a) から (h) はそれぞれ撮影機器 1 から 8 の映像に対する他の撮影機器の映像の類似度を示す。各図に対して、横軸は類似度算出に用いたもう一方の撮影機器番号、縦軸は 2 撮影機器の特徴量のハミング距離である。本評価では、撮影機器同士の番号が近いほど、撮影機器間の距離が近くなることを意味している。

図 9 より、類似度算出に用いた撮影機器同士の距離が近いほど、映像間の類似度を表すハミング距離が小さくなっていることが分かる。マルチビュービデオは複数台の撮影機器が同じ対象物を撮影するため、撮影機器間の距離が近づくほど、映像間の相関が高くなる特性を持つ。各撮影機器の特徴量に基づく類似度算出を用いることで、アクセスポイントによるマルチビュービデオの特性を踏まえた、より正確な送信制御が可能になると言える。

5.3 理想的なトラフィック量との比較

Domino Streaming++ の基本性能を評価するために、映像品質を変化させた場合のトラフィック量を理想的なトラフィック量と比較した。

図 10 に、利用したビデオシーケンスが Ballroom、撮影機器数は 4 台、量子化パラメータを 24 から 40 に変化させた場合の映像ビットレートを示す。横軸は PSNR [dB]、縦軸は映像ビットレート [kbps] である。図 10 より、次の 3 つのことがわかる。

1 つ目は、Domino Streaming++ は、Independent Streaming より映像品質を維持したまま、トラフィック削減を達成していることである。例えば、PSNR が 36 [dB] のとき、Domino Streaming++ は Independent Streaming と比較して、約 250 [kbps] のトラフィック量を削減している。Domino Streaming++ では、他の撮影機器の映像を通信傍受によって取得し、自身の映像のエンコードに利用することで、撮影機器間に存在する冗長な情報を削減している。

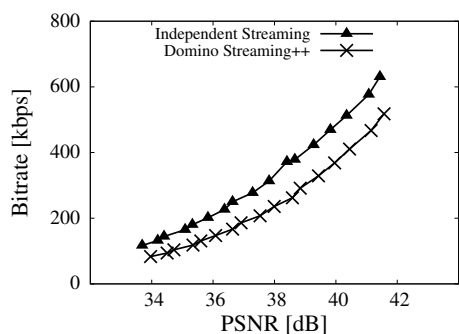


図 12 映像品質に対する映像ビットレート (Exit)

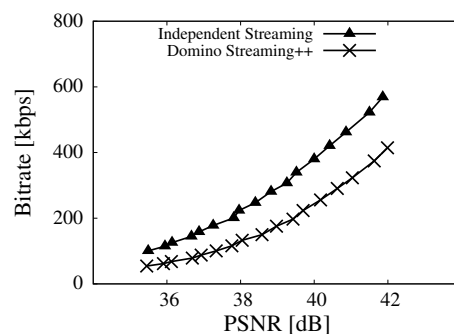


図 13 映像品質に対する映像ビットレート (Vassar)

2つ目は、Domino Streaming++のトラフィック量が、トラフィック量の理想値である Ideal に漸近していることである。例えば、PSNR が 36 [dB] のとき、Domino Streaming++は Ideal と比較して、約 30 [kbps] だけ離れている。Domino Streaming++では、映像類似度に基づく送信順制御を利用することで、撮影機器間に存在する冗長な情報をより多く削減することができ、理想値に近いトラフィック量を達成している。

3つ目は、PSNR が高くなるにつれて、Domino Streaming++と Ideal との差が広がっている点である。例えば、PSNR が 32 [dB] のとき、Domino Streaming++は Ideal と比較して、約 6 [kbps] しか離れていないが、PSNR が 36 [dB] のときは約 30 [kbps] 離れている。PSNR が高くなると、ビデオフレームの低周波成分だけでなく、特徴量には現れない高周波成分にも変化が生じる。高周波成分の変化によって、アクセスポイントが理想的な映像送信順とは異なる映像送信順を決定するため、Domino Streaming++と Ideal とのトラフィック量の差が大きくなっていると考えられる。

5.4 異なるビデオシーケンスにおけるトラフィック量の比較

5.3 節ではビデオシーケンスが Ballroom の場合のトラフィック量を評価した。しかしながら、撮影対象が変化した場合、各撮影機器の映像の特徴量に変化し、Domino Streaming++の性能が変化すると考えられる。このような観点から、異なるビデオシーケンスにおけるトラフィック量を評価した。

図 11 に利用したビデオシーケンスが Ballroom の場合のトラフィック量、図 12 に利用したビデオシーケンスが Exit の場合のトラフィック量、図 13 に利用したビデオシーケンスが Vassar の場合のトラフィック量を示す。それぞれのビデオシーケンスにおいて、撮影機器数は 8 台、量子化パラメータを 24 から 40 に変化させた。横軸は PSNR [dB]、縦軸は映像ビットレート [kbps] である。

図 11 から図 13 より、Domino Streaming++は撮影対象に関係なく Independent Streaming より映像品質を維持した

まま、トラフィック量の削減を達成していることがわかる。例えば、PSNR が 36 [dB] のとき、Domino Streaming++は Independent Streaming と比較して、ビデオシーケンスが Ballroom の場合は約 250 [kbps]、ビデオシーケンスが Exit の場合は約 54 [kbps]、ビデオシーケンスが Vassar の場合は約 57 [kbps] のトラフィック量を削減している。Domino Streaming++では、撮影対象が変化しても、撮影機器間の映像類似度に基づいて送信順を制御することで、撮影機器間に存在する冗長な情報を削減することができる。

6. 終わりに

本稿では、マルチビュービデオの撮影部を無線化する方式として、Domino Streaming++を提案した。Domino Streaming++では、傍受した他の撮影機器の映像を利用して自身の映像をエンコードすることで、トラフィック削減を達成する。また、アクセスポイントが撮影機器間の映像類似度に基づいて、各撮影機器の映像送信順を制御することで、撮影機器間の冗長な情報を多く削減でき、より大幅なトラフィック削減を達成する。性能評価から、提案方式 Domino Streaming++が無線マルチビュービデオにおけるトラフィック量の理想値に近い性能を達成していることが分かった。

参考文献

- [1] Müller, K., Schwarz, H., Marpe, D., Bartnik, C., Bosse, S., Brust, H., Hinz, T., Lakshman, H., Merkle, P., Rhee, H. et al.: 3D High Efficiency Video Coding for Multi-View Video and Depth Data, *IEEE Transactions on Image Processing*, Vol. 22, No. 9, pp. 3366–3378 (2013).
- [2] Tanimoto, M. and Kazuyoshi, S.: Global view and depth (GVD) format for FTV/3DTV, *Three-Dimensional Imaging Visualization And Display*, pp. 1–10 (2013).
- [3] Beck, S., Kunert, A., Kulik, A. and Froehlich, B.: Immersive group-to-group telepresence, *IEEE Transactions on Visualization and Computer Graphics*, Vol. 19, No. 4, pp. 616–625 (2013).
- [4] Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L. and Fuchs, H.: The office of the future: A unified approach to image-based modeling and spatially immersive displays, *ACM GRAPHITE*, pp. 179–188 (1998).

- [5] Vetro, A., Pandit, P., Kimata, H., Smolic, A. and Wang, Y.-K.: *Joint Draft 8.0 on Multi-view Video Coding* (2008).
- [6] Muller, K., Merkle, P., Schwarz, H., Hinz, T., Smolic, A. and Wiegand, T.: Multi-view video coding based on H. 264/AVC using hierarchical B-frames, *IEEE PCS* (2006).
- [7] Text Of ISO/IEC 14496-10:2008/FDAM 1 ISO/IEC JTC1/SC29/WG11: Multiview Video Coding (2008).
- [8] Xiaoyu, X., Gene, C. and Jie, L.: Frame structure optimization for interactive multiview video streaming with bounded network delay, *IEEE ICIP*, pp. 593–596 (2011).
- [9] Ngai-Man, C., Antonio, O. and Gene, C.: Distributed source coding techniques for interactive multiview video streaming, *IEEE PCS*, pp. 1–4 (2009).
- [10] Gene, C., Antonio, O. and Ngai-Man, C.: Interactive streaming of stored multiview video using redundant frame structures, *IEEE Transactions on Image Processing*, Vol. 20, No. 3, pp. 744–761 (2011).
- [11] Liu, Z., Cheung, G. and Ji, Y.: Unified distributed source coding frames for interactive multiview video streaming, *IEEE ICC*, pp. 2048–2053 (2012).
- [12] Huang, H., Zhang, B., Chan, S.-H., Cheung, G. and Frossard, P.: Coding and replication co-design for interactive multiview video streaming, *IEEE INFOCOM*, pp. 2791–2795 (2012).
- [13] Pan, Z., Ikuta, Y., Bandai, M. and Watanabe, T.: User dependent scheme for multi-view video transmission, *IEEE ICC* (2011).
- [14] Pan, Z., Ikuta, Y., Bandai, M. and Watanabe, T.: A user dependent system for multi-view video transmission, *IEEE AINA*, pp. 732–739 (2011).
- [15] Pan, Z., Bandai, M. and Watanabe, T.: A user dependent scheme for multi-view video live streaming, *International Journal of Computational Information Systems*, Vol. 9, No. 4, pp. 1439–1448 (2013).
- [16] Fujihashi, T., Pan, Z. and Watanabe, T.: Traffic Reduction for Multiple Users in Multi-view Video Streaming, *IEEE ICME* (2012).
- [17] Fujihashi, T., Pan, Z. and Watanabe, T.: UMSM: A traffic reduction method on multi-view video streaming for multiple users, *IEEE Transactions on Multimedia*, Vol. 16, No. 2, pp. 1–14 (2014).
- [18] Fujihashi, T., Pan, Z. and Watanabe, T.: Traffic reduction on multi-view video live streaming for multiple users, *IEICE Transactions on Communications*, Vol. 96, No. 7, pp. 2034–2045 (2013).
- [19] Ito, T.: 技研における立体テレビの研究成果, *NHK 技研 R&D*, No. 123 (2010).
- [20] Tanimoto, M.: Overview of free viewpoint television, *Signal Processing: Image Communication*, Vol. 21, No. 6, pp. 454–461 (2006).
- [21] 池谷健佑, 久富健介, 片山美和, 三科智之, 岩館祐一: 多視点ロボットカメラを用いたぐるっとビジョンシステム, *映像情報メディア学会年次大会*, p. 2 (2013).
- [22] Anand, A., Akella, A., Sekar, V. and Seshan, S.: A case for information-bound referencing, *ACM SIGCOMM*, p. 4 (2010).
- [23] Anand, A., Balachandran, A., Akella, A., Sekar, V. and Seshan, S.: Enhancing video accessibility and availability using information-bound references, *ACM SIGCOMM*, pp. 345–356 (2013).
- [24] Lowe, D. G.: Object recognition from local scale-invariant features, *IEEE ICCV*, Vol. 2, pp. 1150–1157 (1999).
- [25] Xun, G., Yan, L., Feng, W., Wen, G. and Shipeng, L.: Distributed multi-view video coding, *VCIP*, Vol. 38, No. 11, pp. 1917–1921 (2006).
- [26] Artigas, X., Angeli, E. and Torres, L.: Side information generation for multiview distributed video coding using a fusion approach, *IEEE NORISIG*, pp. 250–253 (2006).
- [27] Frederic, D., Mourad, O. and Touradj, E.: Recent advances in multiview distributed video coding, *DSS*, pp. 1–11 (2007).
- [28] Slepian, D. and Wolf, J. K.: Noiseless coding of correlated information sources, *IEEE Transactions on Information Theory*, Vol. 19, pp. 471–480 (1973).
- [29] Wyner, A. and Ziv, J.: The rate-distortion function for source coding with side information at the decoder, *IEEE Transaction on Information Theory*, Vol. 3, No. 4, pp. 45–49 (1976).
- [30] IEEE Computer Society: *IEEE Standard for Information technology-Telecommunications and information exchange between systems Local and metropolitan area networks-Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications* (2012).
- [31] Joint Video Team Of ITU-T VCEG And ISO/IEC MPEG: *JMVC (Joint Multiview Video Coding) Software* (2008).
- [32] ISO/IEC JTC1/SC29/WG11: *Multiview Video Test Sequences from MERL* (2005).