

小型撮影機器を用いたマルチビュービデオ伝送の基礎方式について

小寺志保[†]藤橋卓也[‡]猿渡俊介[†]渡辺尚[‡][†] 静岡大学[‡] 大阪大学

1 はじめに

マルチビュービデオは、臨場感のある映像を提供する一技術として注目を浴びている。具体的なアプリケーションとして、立体映像 [1]、自由視点映像 [2]、テレプレゼンス [3] が挙げられる。マルチビュービデオ技術は、これらのアプリケーションを誰でも気軽に利用可能となることを目標としている。

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。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、撮影部を無線化し、場所を選ばなくマルチビュービデオが撮影できるシステムを考える。マルチビュービデオの撮影部の無線化を実現する方式として、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フィックを削減する。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、アクセスポイントによって映像送信順を制御する。

2 要件

図 2 に無線化した撮影部を示す。各撮影機器からアクセスポイントまで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に対して伝送する。マルチビュービデオの撮影部を無線化する場合、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、トラフィックを低くするという点である。本研究では無線通信で映像を伝送することを想定しているため、帯域が有線ほどには大きくない。マルチビュービデオは複数のカメラからの映像を扱うことに起因して、シングルビュービデオよりトラフィックが大きくなる。少ない帯域に対してトラフィックが大きい場合には伝送に時間がかかり、ユーザが映像を切り替えても再生まで時間がかかるなどの問題が発生する。

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、元の映像と比較して劣化が小さい映像を指す。視聴者は、現実世界と同等の品質の映像を求める。3D 映像を作成することを考えても、高品質な映像の方がより高精細な 3D 映像を作成することができる。

しかしながら、映像の品質とトラフィックはトレードオフの関係にある。一般的に、スムーズな視聴のために低トラフィックを実現しようとすると映像の品質は低下し、高い映像品質を実現しようとするとトラフィックは高くなる。

3 Domino Streaming

本研究では、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 は、初期化、送信順決定、エンコード、映像伝送、デコードの 5 つのフェーズで構成される。

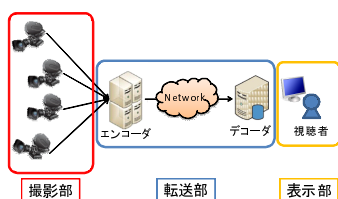


図 1: マルチビュービデオシステムの構成

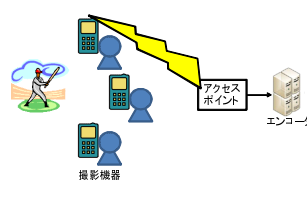


図 2: 無線化した撮影部

3.1 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各撮影機器に ID を割り当てるため、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。アクセスポイントの通信範囲に入った撮影機器はビーコンを受信すると、ID を取得するために Information-Bound References の値を含んだプローブ要求フレームをアクセスポイントに対して送信する。

Information-Bound References (IBR) [4, 5] とは、各撮影機器が撮影した映像の特徴量を算出する方法である。アクセスポイントは各撮影機器が算出した IBR 値を用いることで、各撮影機器から映像を受信することなく、撮影機器間の映像の類似度を求めることができる。IBR 値の算出には、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。以下に、各撮影機器で IBR 値を算出する方法を示す。

まず、各撮影機器は自身が撮影した映像の 1 GOP 分のビデオフレームのうち、先頭のビデオフレームの解像度を 128×128 画素に変換する。GOP とは、複数のビデオフレームの集合体のことであり、通常は 8 枚のビデオフレームから構成される。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。撮影機器は続けて輝度信号 Y、色差信号 Cb, Cr に対して DCT 変換を実行し、各信号の低周波成分と高周波成分を取得する。撮影機器は取得した周波数成分を用いて、IBR 値を算出する。

IBR 値は 4 つの要素から構成される。1 つ目は、 128×128 画素の輝度信号 Y のうち、先頭の 8×8 画素の低周波成分から算出する。具体的には、まず撮影機器は合計 64 個の輝度信号 Y の周波数成分から中央値を算出する。その後、各周波数成分が中央値を超えているかに応じて、1 か 0 を割り当てる。撮影機器は最終的に得られる 64 個の 0, 1 列を IBR 値の 1 要素として用いる。2 つ目は、輝度信号 Y の残りの高周波成分の和である。3 つ目は、 128×128 画素の色差信号 Cb のうち、先頭の 3×3 画素の低周波成分である。4 つ目は、 128×128 画素の色差信号 Cr のうち、先頭の 3×3 画素の低周波成分である。

アクセスポイントは撮影機器から IBR 値を含んだプローブ要求フレームを受信すると、撮影機器に対して ID を割り当てフレームを送信し、撮影機器に ID を割り当てる。

3.2 送信順決定

アクセスポイントは撮影機器間にある冗長な情報を効率的に削減するため、各撮影機器から受信したプローブ要求フレームに含まれる IBR 値を元に、各撮影機器の映像送信順を決定する。IBR 値を用いることで、各撮影機器がエンコード時に自身の映像と類似した他の撮影機器の映像を参照できるようになる。類似した映像を参照してエンコードすることで、撮影機器間にある冗長な情報をより多く削

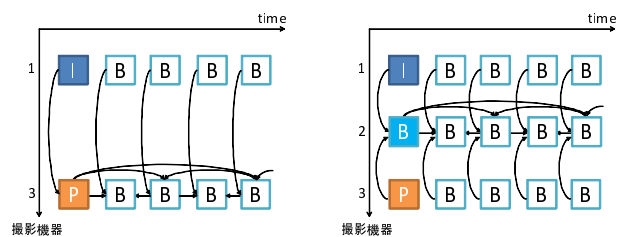


図 3: 撮影機器 3 の映像のエンコード 図 4: 撮影機器 2 の映像のエンコード

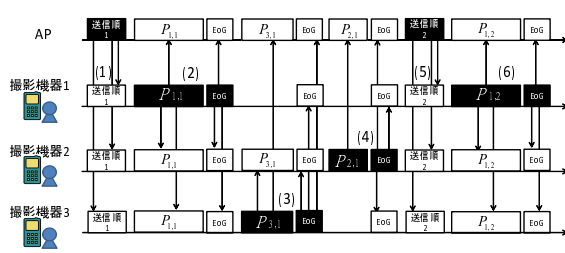


図 5: 映像伝送タイムシーケンスチャート

減できる。さらに冗長な情報を削減するために、アクセスポイントは送信順決定時に、類似した映像を持つ撮影機器間で H.264/AVC における双方向予測が利用可能か確認する。双方向予測とは、ある撮影機器が自身の映像をエンコードするとき、自身の前後または左右の撮影機器が撮影した映像を利用することで、より大幅なトラヒック削減を達成する技術である。双方向予測が利用できる場合は、アクセスポイントが IBR 値による映像の類似度と双方向予測を考慮した映像送信順を決定する。

映像送信順が決定すると、アクセスポイントは全ての撮影機器に決定した送信順をブロードキャストする。

3.3 エンコード

3.2 節の方法で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 (Group Of Picture) 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ヒ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。

例として、アクセスポイントの通信範囲内に撮影機器 1, 2, 3 があり、アクセスポイントが撮影機器 1, 3, 2 の順に送信順を決定したと仮定する。撮影機器 1 の映像は単独でエンコードされるため、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 3 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 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 3 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 フレームを基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 4 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 と 3 の映像を傍受しているため、撮影機器 1, 3 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 2 の先頭フレームを B フレームとしてエンコードする。B フレームは複数の撮影機器の映像を基に作成した自身の映像との差分情報であるため、最もデータ量が少なくなる。

3.4 映像伝送

図 5 に映像伝送のタイムシーケンスチャートを示す。図 5 では、アクセスポイントの通信範囲内に撮影機器 1, 2, 3 があると仮定し、送信順は撮影機器 1, 3, 2 の順とする。 $P_{i,j}$ 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、各撮影機器でエンコードした映像と、3.1 節で述べた IBR 値を含んでいる。映像の変化に対応するために、各撮影機器は 1 GOP ごとに IBR 値を求める。

1. アクセスポイントは 3.2 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ブロードキャストする。
2. 撮影機器 1 がアクセスポイントに対して $P_{1,1}$ を伝送する。撮影機器 2 と 3 は $P_{1,1}$ を傍受し、撮影機器 1 の映像をデコードする。 $P_{1,1}$ を送信し終わった撮影機器 1 は、EoG (End of GOP) パケットを送信する。EoG パケットは、1 つの GOP を送信し終わったことを意味する。
3. $P_{1,1}$ と撮影機器 1 の送信した EoG パケットを傍受した撮影機器 3 は、3.3 節の方法にもとづいて撮影機器 3 の映像をエンコードした後、 $P_{3,1}$ を送信する。撮影機器 2 は $P_{3,1}$ を傍受し、撮影機器 3 の映像をデコードする。 $P_{3,1}$ を送信し終わった撮影機器 3 は EoG パケットを送信する。
4. $P_{1,1}$, $P_{3,1}$, 撮影機器 3 の送信した EoG を受信した撮影機器 2 は、3.3 節の方法にもとづいて撮影機器 2 の映像をエンコー

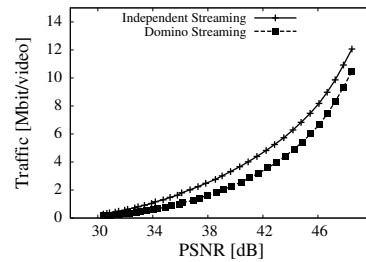


図 6: 映像品質に対するトラヒック量

ドし、 $P_{2,1}$ を送信する。 $P_{2,1}$ を送信し終わった撮影機器 2 は EoG パケットを送信する。

5. 撮影機器 2 の EoG パケットを受け取ったアクセスポイントは、各撮影機器の IBR 値から次の GOP を送信するための送信順を決定する。
6. アクセスポイントがブロードキャストした送信順を受信すると、送信順に従って次の GOP を送信する。

Domino Streaming ではすべての GOP の送信が終了するまで、以上の動作を繰り返す。

3.5 デコード

デコードには標準的な H.264/AVC のデコーダを利用する。各撮影機器が最初に傍受する映像は他の撮影機器の映像を参照せずエンコードされているため、各撮影機器が独自にデコードする。続けて傍受する映像は、すでにデコードし終えた映像を元に作成された差分情報であるため、各撮影機器はデコードを終えた映像を用いて受信と同時にデコードする。

4 性能評価

位置を固定した撮影機器 8 台における、映像品質に対するトラヒック量を測定した。JMVC エンコーダ [6] と MERL が提供しているテストビデオシーケンス [7] の Ballroom を用いた。1 GOP あたりのフレーム数は 8 フレームとし、量子化パラメータを 9~50 に変化させた。比較対象として、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 Independent Streaming を用いた。図 6 に、映像品質に対するトラヒック量を示す。図 6 より、Domino Streaming は、Independent Streaming より映像品質を維持したまま、トラヒック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、約 700 [Kbits/video] のトラヒックを削減している。

5 おわりに

本稿では、マルチビュービデオ撮影部の無線化を実現するための伝送方式として、Domino Streaming を提案した。Domino Streaming は、各撮影機器が他の撮影機器の通信を傍受し、エンコードに利用することでトラヒックを削減する。さらに、トラヒック削減効果を高めるため、アクセスポイントが映像の類似度を元に、各撮影機器の送信順を制御する。性能評価から、Domino Streaming が既存方式と比較して、映像品質の劣化を防ぎながら、トラヒック削減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] Anand, A., Akella, A., Sekar, V. and Seshan, S.: A case for information-bound referencing, *ACM SIGCOMM*, p. 4 (2010).
- [5] 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).
- [6] Joint Video Team Of ITU-T VCEG And ISO/IEC MPEG: *JMVC (Joint Multiview Video Coding) Software* (2008).
- [7] ISO/IEC JTC1/SC29/WG11: *Multiview Video Test Sequences from MERL* (2005).