



テレワークに有効な 映像メディア処理に関する取り組み

西 敬

OKIでは、次世代テレワークシステムの研究開発を行っている。互いの場所が離れていたとしても、臨場感あふれるコミュニケーションをとることで、知識労働者の生産性を向上させることができると考え、日々研究を行っている。

究極の臨場感コミュニケーションは、「以心伝心」であると考えている。何も言わなくても、心と心が通じていることが理想である。以心伝心のためにキーとなるのは、コミュニケーション相手自身の理解、状況の理解、気持ちの理解であると考えている。それらの理解により、コミュニケーションに必要な臨場感が感じられると考えている。

その上で映像メディアがすべきことは、理解を助ける、促進するための情報を伝達することだと考えている。何が理解を助ける情報なのか、何が理解を促進するための情報なのかを見つけることが重要な研究課題であると同時に、それらの情報を映像技術によりどのように伝達するかも重要である。現在我々は、それらの情報を伝達する重要な技術の候補として、以下の技術に対して研究を行っている。

(1) 超低演算量な映像符号化方式を用いた多視点映像通信技術

(2) 映像加工・強調技術

以下では、これら二つの技術について、我々の現在の取り組みを紹介する。

常時接続実験による新たな課題の発見

現在我々は、大阪のオフィスと藤のオフィスに別れ、テレワークを行っている。既存のテレビ会議システムを常時接続し、日々のコミュニケーションに与える影響を調査している。コミュニケーションを円滑にするには、コミュニケーションをとるときだけのつながりではなく、普段のつながりが重要であると考えているからである。実際に使ってみると、遠隔地にいる研究員を（個人的には）身近に感じることができる。もちろん普段からメールや

電話のやりとりなどはしているが、やはり「百聞は一見にしかず」であることを改めて感じる。

日々の運用で最も効果を感じたのは、「プレゼンス」を確認してから電話できることである。この場合のプレゼンスとは、電話をかける前に相手が着席しているだけでなく、電話をかけても良い状態であるかどうかまで含めたプレゼンスである。

常時接続によりさまざまな効果が感じられるが、以下のような課題も発見された。

- 着席していないときに、どこに行っているか、もしくは社内には不在なのかわからない。
- 話をしている相手の表情が分かりにくい。
- オフィス全体の雰囲気分かりにくい。
- 自分の声が相手にどのように聞こえているかが分かりにくい。

課題を整理すると、次の3点が重要であることに気付いた。

- ①オフィスの全体の「今」の雰囲気を伝えること
 - ②コミュニケーション相手の状況や気持ちを伝えること
 - ③「普段」のオフィスの状況を分かりやすく伝えること
- ①、②を実現するためには、カメラ1台での実現は困難であることに気づき、多視点映像に着目し技術開発を進めている。

また、③を実現するためには、「映像加工・強調技術」が有効であると考えている。

以下では、これらの技術が有効であると考えている理由と、現在の研究開発状況を説明する。

多視点映像通信への取り組み

我々は、オフィス全体の雰囲気を伝えるためには、「オフィス全体の映像」と「コミュニケーションを頻繁にとる人の現状を示す映像」の両方を提示する、すなわち多視点映像を用いることが有効であると考えている。たとえば、「着席していないときに、どこに行っているか、も

しくは社内には不在なのかがわからない」という課題に対しては、着席をしていないことがわかる俯瞰映像と行き先表示板の映像などをセットにして提供することなどがある。

一方、コミュニケーション相手の状況や気持ちを伝えるためには、「相手の顔や手足の動きなどを含めた映像を提示すること」が有効であると考えている。たとえば、話をしている相手の表情が分かりやすくなるようにクローズアップすると同時に、手や足の映像も提示することで、必ずしも表情には出てきていない感情を伝えることができる。

また、複数の映像から自由視点映像を作成したりすることで、より実際の状況に合った映像表示を行うことなども有効であると考えている。

このように多視点映像は有効であると考えられるが、符号化時に多くの演算量が必要であるという課題がある。符号化の演算量に関しては、視点が複数になる（カメラが複数、入力画像が複数と同意）ことから、単純に考えて“視点数”分の符号化処理が必要になる。現在映像符号化の主流となっているH.264/MPEG-4 AVC¹⁾などは、フレーム間符号化（過去の画像との差分を符号化）の際に行う動き補償技術により圧縮率が高い反面、演算量が多いことで知られている。また、多視点映像の圧縮率をさらに向上させる方式がH.264/MPEG-4 AVCの拡張方式（Multi-view Video Coding²⁾）として規格化されようとしているが、フレーム間の動き補償技術に加え、視点間の補償技術などを利用するため、さらに多くの演算量が必要である。

そこで我々は、符号化処理を軽減できる新しい映像符

号化方式であるDistributed Video Coding³⁾（以下DVC）に着目し、リアルタイムな多視点映像通信システムを実現するべく研究を開始している。DVCとは、「相関を有する二つのデータを独立に符号化しても、参照して符号化したときと同様に効率的に圧縮できる」という理論（Slepian-Wolf理論、Wyner-Ziv理論）を動画像符号化に応用した方式である。

DVCの符号化の特徴は、誤り訂正符号化を用いているところにあり、キーフレームと非キーフレームは異なる符号化方式で符号化する（図1）。キーフレームは、たとえば従来から利用されているフレーム内符号化を行う。一方、非キーフレームに対しては誤り訂正符号化を行い、その結果計算された検査ビットのみを受信側に送る。復号側では次フレームの画像を予め推測し（たとえば過去2フレームの画像から推測する）、その推測画像と次フレームの画像との差分（推測誤り）を送られてきた検査ビットを用いて検査および訂正することで次フレームの画像を得ることができる。このようにDVCは符号化側では動きベクトル探索は行わずに符号化するため、一般の符号化方式と比較して、条件により異なるが1/50～1/200の演算量で符号化できる。欧米では活発に研究が行われており、特にヨーロッパの次元プロジェクトDISCOVERなどは、個別の技術の研究だけでなく、プラットフォームの研究⁴⁾も行っている。

DVCは、符号化効率という面で従来の動画像符号化方式との差はまだ大きいですが、符号化の演算量が従来と比較して極端に少ないことから、ワイヤレス環境での監視装置、携帯電話などを用いたテレビ電話に期待されている

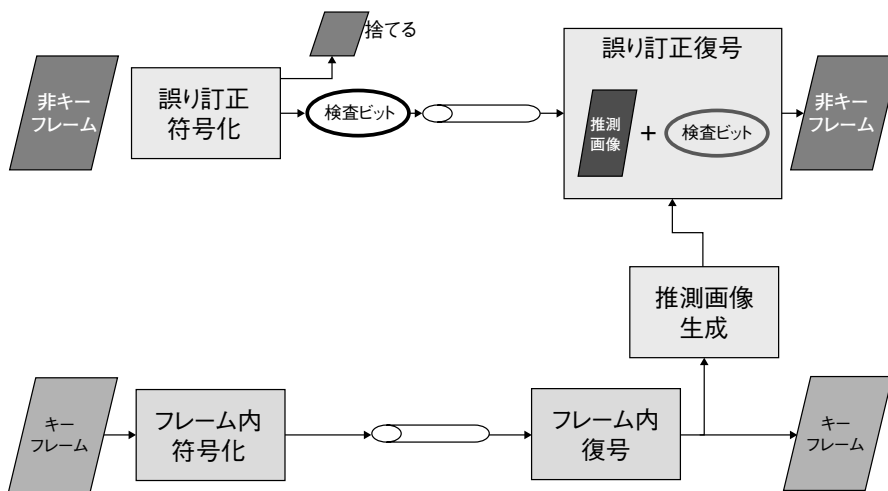


図1 DVCの符号化方式の基本構成

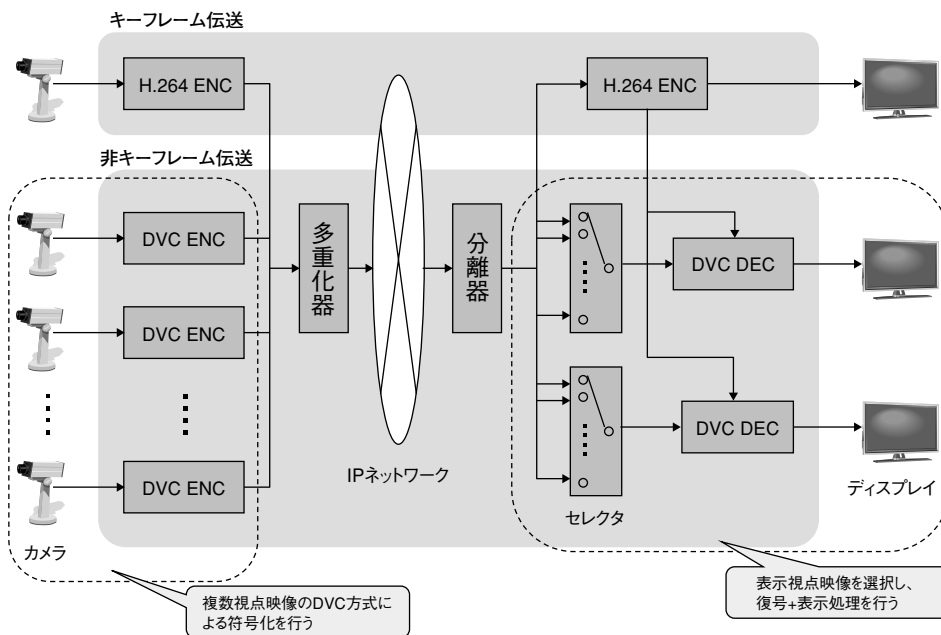


図2 リアルタイム多視点映像通信システムの構成

と同時に、今回我々が着目している多視点映像システムへの応用も考えられている。

次に、我々が実現しようとしているリアルタイム多視点映像システムの構成例を図2に示す。このシステムは欧米で主流となっているDVCプラットフォームをベースにリアルタイムに動作する映像通信システムを開発/実現するものである。送信側には複数のカメラ（たとえば5視点）の映像のうち4視点の映像をDVCにより符号化/伝送し、受信側では5視点の映像から3視点の映像（この例ではキーフレームは常時表示）を選択し、DVC復号処理および表示をリアルタイムに行うシステムである。

このシステムを実現するために、以下の3つの技術課題に着目し、研究を進めている。

①フィードバックの不要な符号量制御技術の開発

現在DVCの研究で主に行われているのは、復号器で復号した後、その復号結果が十分であるかを判定し、不足がある場合には、符号化器に追加の情報を送信するよう要求する、いわゆるフィードバックを必要とする方式である。この方式は、復号器の演算量を増加させるだけでなく、遅延を大きく増加させることになり、リアルタイム処理には適さない。我々はこの問題を解決するために、フィードバックを必要としない符号量制御を開発する。

②復号器によるフレーム間相関の推測技術の開発

DVCの復号の際には、フレーム間の相関が必要となる。

現在のDVCの研究では、フレーム間の相関が既知であると仮定しているが、リアルタイム環境で動作させることは不可能である。我々は、復号器でフレーム間の相関を高精度で推測する技術を開発し、リアルタイム処理を実現する。

③復号器の並列処理アルゴリズムの開発

DVCの復号処理は既存の映像の復号処理と比較して、非常に多くの演算量がかかる。それはDVCの特徴である誤り訂正符号（我々が目指しているシステムでは、LDPC符号を用いる）の復号処理に多くの演算量がかかるからである。我々は復号処理を並列に行うことや、検査ビットだけでなく情報ビットも適応的に送信することで、リアルタイム処理を実現する。

映像加工・強調技術への取り組み

我々は、「普段のオフィスの状況を分かりやすく伝えること」が重要であると考えている。

普段のオフィスの状況とは、「いつも同じ空間で働いていれば伝わりやすいが、テレワークの場合は伝わりにくい情報」である。

たとえば、

- なんだか最近忙しそう
- いつもこの部署は盛り上がっているなあ
- 何か面白そうな実験をいつもやっているなあ

●今日は職場全体が夕方に帰った（飲み会だった）
などである。

このような状態を知るのに、映像をじっと見続けるというのは現実的ではないのは容易に推測できる。よって、「見るとはなしに見る」ことで、普段の状況をだまかに把握することが必要であると考えている。たとえば何か特徴的なことが起こったときにその状況を映像処理により強調することが有効であると考えている。

図3に示したように「盛り上がっている、人の声が入っている、笑い声が入っているなどのときは、その一帯をほんのり黄色、そうでないところはうすい水色で表示し、人通りが多いときにはほんのり赤く表示する」が一例である。



図3 職場の状況の強調処理の一例

他に、テレワークの場合に伝わりにくい情報として、「自分の声が相手にどのように聞こえているかが伝わりにくい。」ということもある。自分の声がどのように伝わっているかわからないために、必要以上に大きな声になったり、また、想像以上に遠くの人にまで、自分の声が聞こえてしまったりすると、話しにくくなり、人と人とのコミュニケーションの基本である「会話」がスムーズにできない。そのようなときは、自分の声がどこまで、どのように届いているかを一目で分かるようにすることで、話しやすくなる。そのため、コミュニケーションが活発になり、相互理解も促進される。

図4に示したように、「自分の声が届いている範囲とそうでない範囲の明るさを区別して表示する」が一例である。

状況理解という領域は人間がどう感じるかに依存するので、実際に検証実験が必要である。検証すべきテーマはたくさんあるが、我々が現在取り組んでいる一例として、映像加工・強調技術を紹介した。

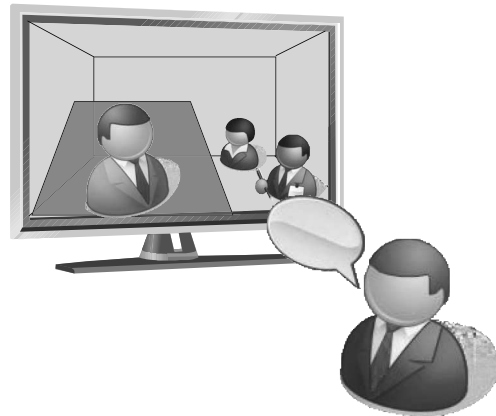


図4 自分の声の伝わり状況を示す一例

ま と め

臨場感コミュニケーションを実現する映像メディア処理技術として、我々が着目した課題、要望、解決し得る映像技術と我々の現在の取り組みを紹介した。

あるWorkshopで「コミュニケーションの質を決めるのは情報の発信者ではなく、受信者である。」という話を聞いたことがある。もちろん情報の発信者も重要であるのは言うまでもないが、実際に決めているのは受信者であるという意味だと思われる。受信者がどのように決めているかの大きな要因の一つは、冒頭でも述べたようにさまざまな観点からの「理解」であると考えている。今まで実現されているシステムでは、遠隔地における「受信者」が「理解」をするための仕組みを十分提供していないと考えている。我々はこれからも、「理解を促進する映像処理」という観点にこだわり、技術開発を行っていく。



参考文献

- 1) ISO/IEC 14496-10:2005, Advanced Video Coding (Third edition), December, 2005
- 2) "Study of Text of ISO/IEC 14496-10: 200X/FPDAM 1 Multiview Video Coding", ISO/IEC JTC1/SC29/WG11, N9760, Archamps, France, April, 2008
- 3) B.Girod, *et al.*: "Distributed video coding," Proceedings of the IEEE, Special Issue on Video Coding and Delivery, vol.93, no.1, pp.71-83, January 2005. Invited paper.
- 4) X. Artigas, *et al.*: "The DISCOVER CODEC: Architecture, techniques and Evaluation", PCS2007, Nov. 2007

● 筆者紹介

西 敬 : Takashi Nishi. 研究開発本部 ヒューマンコミュニケーションラボラトリ