

多地点TV会議システムにおける コヒーレントな過去情景再生

青柳 滋己¹ 高田 敏弘¹

概要:

今のビジネスシーンにおいて、会議や共同作業等、遠隔地とのリアルタイムコミュニケーションは欠かすことのできない重要なものとなっている。また、カメラデバイスや映像・音声等の蓄積デバイスも高速・大容量となり、会議を高画質で丸ごと録画しておき、後から参照するといったことも容易にできるようになっている。一方、蓄積した会議等の情報は、まだ有効活用できていないと言いがたい。我々は時空間同室感システム t-Room の設計・開発を行っている。t-Room では空間配置の整合性を保持したまま録画・再生することが可能で、複数遠隔地での会議や作業をコヒーレントに再生する。本稿では t-Room 録画再生システムの詳細と応用例について述べる。

Coherent playbacks of past activities on a multilocal video communication system

SHIGEMI AOYAGI¹ TOSHIHIRO TAKADA¹

Abstract: Video communication systems are now popular in the business scene. It helps talking with people in distant sites. Some systems have a record/playback functions of conversation in the systems. However, recorded data cannot be used efficiently. We have designed and developed a system called t-Room. In this paper, we describes the details of coherent record/playback functions on a multilocal t-Room system.

1. はじめに

TV 会議システムは遠隔地との円滑なコミュニケーションのためによく使われており、ビジネス分野においては今や必須システムとなっている。ネットワークや画像の高解像度化に伴い、遠隔地の相手の様子が臨場感を伴って鮮明に、違和感少なくコミュニケーションできる環境が整いつつある。また、ノート PC やタブレット PC、インターネット技術の発達により、場所による制限も少なくなり、好きな時に好きな場所で会話や会議を行うことも可能となっている。TV 会議システムの中には、録画・録音機能を提供しているものがあり、過去に行った会議の様子を再生して見ることが可能である。映像を蓄積するための HDD といった記憶媒体も安価・大容量化し、映像・音声を 1 会議分全

部録画することも可能である。

しかし、従来のシステムにおける過去録画・再生は、システム内の様子を単に記録・再生して見ることが可能なだけであり、過去映像を今の会議を行っている最中に積極的に利用するようことはできない。つまり過去の録画・再生は、ただ単に見ることができただけで、それ以上の活用がしにくい。

我々は、遠隔会議・コラボレーションシステムである t-Room[1][2] の研究開発を行っている。t-Room システムでは録画と再生が同じシステム上で実現できるため、時間的に離れた過去を、距離的に離れた地点を映すのと同様に空間配置において矛盾のないコヒーレントな状態で再生表示が可能である。この再生機能により、過去の映像を積極的に使用することができる。

本稿では t-Room のコヒーレントな映像再生機能のモデルとその実装について述べる。

¹ NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories

以下、2章ではt-Roomシステムの概要について、3章では関連研究について、4章ではコヒーレントな過去情景再生について、5章では実装について述べ、6章は議論とする。

2. t-Room システム

t-Room は NTT コミュニケーション科学基礎研究所で研究開発されてきた遠隔コミュニケーション支援システムである (図 1)。t-Room の狙いは遠隔コミュニケーションにおける距離と時間の壁を超えることにある。t-Room システムではディスプレイに囲まれた空間を作り、中の様子をカメラで撮影して遠隔地の人や物をその部屋の壁に映すことにより、あたかも同じ部屋にいるような同室感を作り出す。現在のシステムは 65inch 縦置き液晶ディスプレイを 8 枚使い、10 角形の 8 面に配置して 1 部屋を構成している。一般の TV 会議システムでは、ディスプレイの正面に座った人をディスプレイの上に設置されたカメラが撮影し、遠隔地に映像を送る。一方、t-Room では一般の TV 会議シ



図 1 t-Room システム概観

ステムとは異なり、人はディスプレイのすぐ前にディスプレイに背を向ける向きで立ち、ディスプレイの正面にカメラを設置し、ディスプレイの直前に立つ人を撮影してその映像を遠隔地の同じ位置のディスプレイに等身大で表示する (図 2)。カメラは人の背後にあるディスプレイも撮影してしまうと合わせ鏡のように映像が発振してしまう。それをキャンセルするために、カメラのレンズ前に偏光フィルタを付け、ディスプレイ映像は撮影せずにその前の人の映像のみを撮影している。遠隔地の人は、遠隔地でその人が立っている位置に相当するディスプレイに映るので、部屋の中では相手との位置関係はどちらの地点でも同じになり、空間配置の整合性が保たれる (コヒーレントな状態)。

ディスプレイには相手の映像だけでなく、写真、動画像、PC の画面などを任意の位置に重ねて表示できるため、同じ写真や資料を見て会議可能であり、また相手の指さしの

位置なども伝えることができる。また、3 地点、4 地点と接続した場合でも、すべての地点で遠隔地の人との位置関係が同じになる。現在、同じ構成のシステムが京都と厚木に 2 セットずつ設置されている。また、t-Room ではテーブルも設置可能である。50inch 液晶ディスプレイ 2 台で構成されたテーブルと、それを天井から撮影するカメラも設置しており、テーブル上の書類などを共有することが可能である [3]。

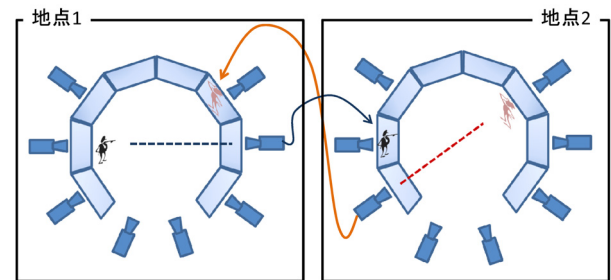


図 2 カメラとディスプレイの配置

3. 関連研究

3.1 ビデオコミュニケーション環境

ビデオカメラ/ディスプレイとネットワークを用いたテレコミュニケーション環境の研究は古くから行われている。代表的なものは VideoWindow[4] であり、面の向こうに相手の空間が映るという一般のテレビ電話である。一方、VideoDraw[5] はディスプレイの表面が共有面となっている。同室感通信 [6], Agora[7], t-Room は VideoDraw のような共有面を組み合わせることで共有空間を実現している。

現在販売されている大型 TV 会議システムには高精細化や低遅延化を重視した製品が多い。複数台のカメラとディスプレイを使用したシステムもあり、それらはテーブルとディスプレイの配置を工夫し、一つのテーブルの向こう側とこちら側に座って話をしているように遠隔地の人を映すシステム、更に音声もその位置から聞こえるように工夫した高臨場感をうたったシステムもある。また、それらのシステムではカメラや音声の録画機能を備えたものもある。

HyperMirror[8] では相手の空間を重畳し、鏡の世界 (仮想空間) 上で相手との共有空間を実現している。仮想空間内の位置関係はすべての地点で同じに保たれる。また、この方式では過去映像の再生もコヒーレントに行うことが可能である。t-Room では同じ地点にいる人は通常の会議と同じ面と向かって会話できるが、HyperMirror では同じ部屋の人とも面と向かうのではなく仮想空間を見て会話する点が異なる。

PC を用いた Web 会議システムやチャットシステムでは、標準的に多地点接続を行うことが可能である。しかしながら、これらの PC を使った TV 会議システムは画面も上半身だけ映し映像自体が小さい場合が多い。

3.2 ビデオコミュニケーションの記録と再生

ビデオコミュニケーションシステム (VCS) 研究の初期の段階から、システム内で行われた活動の様子をビデオとして記録し、さらにそれを後から参照することの有効性についての論が活発にあった [9]。以下、TV 会議システム等の記録と再生に関する関連研究を挙げる。

Media Space [10] は遠隔地間を結ぶ VCS であり、システム上で行われたミーティングは常に録画され、欠席者等が後に録画を再生することでミーティングの内容を把握することができる。Reactive Room [11] や文献 [12] のシステムは、発表者や聴講者を写す複数のビデオカメラ、共有ホワイトボード、コンピュータ等の出力を遠隔地に送ると同時に、全ての映像を録画し、ミーティングの様子を記録する。これらのシステムは活動の記録はできるものの、その再生視聴はビデオモニターや Web ブラウザ等の会議システム上で行われるため、人や物の位置関係・配置関係を保ったまま再生することは不可能である

Reflection of Presence [13] は利用者の前に置かれる鏡像タイプの単一ビデオ共有面からなる VCS である。ビデオ共有面上で録画されたデータの再生が可能であり、またリアルタイムのコミュニケーションとの混在も可能である。VideoPassage [14] は単一ビデオ共有面を記録再生装置とするビデオメールシステムである。記録されたビデオメッセージをスクリーンに表示再生し、そのビデオに対して引用や加筆をしながら返答する様子を録画し、それを新たなメッセージとして記録する。これらのシステムは、記録と再生が同一システム上ででき、人や物の位置関係もある程度保たれるが、どちらも 1 枚の画面で構成されるシステムであり、空間全体の記録再生には至っていない。

4. コヒーレントな過去情景記録再生

通常、遠隔地間の TV 会議の記録映像を再生する際、図 3 のようにディスプレイに表示するため、会議をしている遠隔地と過去映像の位置関係が保たれなくなる場合がある。さらに、過去の映像で参照していた過去映像は画面上に小さく映し出されることになり、再生している過去は何を見て議論していたのかわからなくなってしまう問題がある。

我々が提案するコヒーレントな記録・再生とは、次を満たすものである。

- (1) システム内の人、物、資料といった映像や会話音声だけでなく、配置関係も含めて記録することができる。
- (2) 映像中の人や物の位置関係を保ったまま再生することができる
- (3) 多地点接続中の過去再生は、接続しているすべての地点で同じ映像を見ることが可能
- (4) 再生される過去映像中の人や物と、それを見ている現在の人や物との位置関係は、すべての地点で同じ
- (5) 過去映像を見ている過去を再生した場合も、過去映像

中に参照している映像が確認できる

1,2 については、t-Room では部屋を丸ごと録画しているため、そのまま再生すると位置の情報はそのまま記録されることになる (つまりカメラ配置により、位置情報も記録できている)。会議や会話ではどの人がどこに居て何を話していたのかを保ったまま再生できる。

3 については、t-Room では位置関係はリアルタイムの通信では完全に保たれる。また、その過去情景を録画し、そのまま再生すれば過去に録画した位置関係も保たれたままである。したがって、4 も同じように保たれる。

5 についてはディスプレイの一部、あるいは専用ディスプレイを用いて過去の情景を再生する方式の場合、図 3 のように過去再生している様子を録画し再生すると再生している画面の縮尺がどんどん小さくなってしまう。録画サイズと再生画面のサイズが一致している t-Room では画面の縮尺が変わらないため、再生サイズによる画面劣化は起こらない。

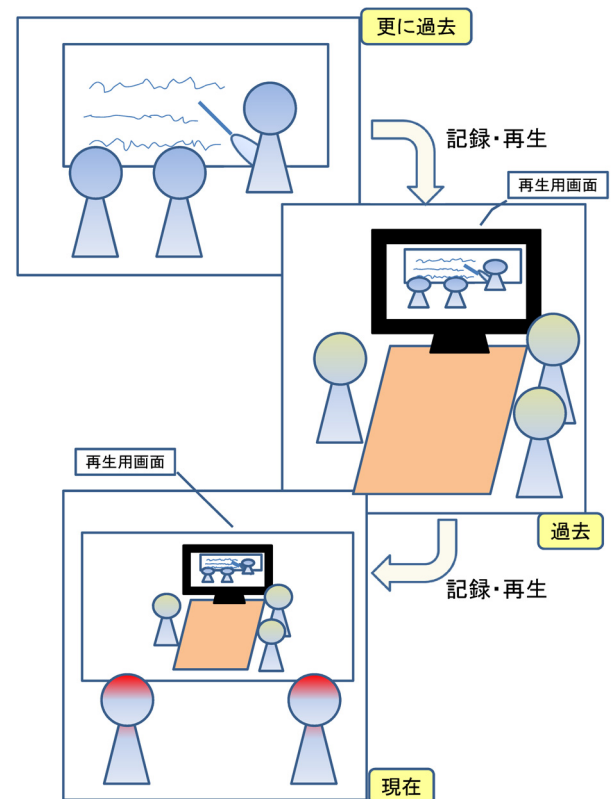


図 3 コヒーレントでない過去再生

5. 実装

5.1 映像記録システムの構成

現在の t-Room は、ディスプレイ 1 枚ごとに 1 台の PC を使用し、また、HD カメラから映像をキャプチャしネットワークに送るため、カメラ 1 台につき 1 台の PC を使用している。1 地点の t-Room は 8 枚のディスプレイ、カメ

ラから構成され、他に記録用 PC1 台、制御用 PC が 1 台の全 18 台構成である。

t-Room のソフトウェアは、カメラからの映像をキャプチャし、ネットワークを通して相手に送るキャプチャ機能、ディスプレイ上にカメラからの映像や静止画等を映し出す表示機能、映像や音声を記録・再生したり画像サイズを変更したりする proxy 機能の 3 つから成る。Windows XP, 7 上で動作し、単一のバイナリが起動時の引数の切り替えにより、上記 3 つの機能を使い分けることができる。

映像や音声の記録には proxy 機能を用いている。proxy 機能は

- 表示機能と同様にキャプチャ機能からデータを受け取る
- キャプチャ機能と同様に表示機能にデータを送る
- 画像データのサイズの変換や複数映像を重ねて 1 つの映像に合成する

という機能を同時に行うことが可能である。近年、カメラの高精細化の進歩が著しく保存すべきデータ量が増加しつつあり、キャプチャデータの保存のために大容量・高速な保管デバイスが必要となっている。この proxy 機能により、キャプチャと録画する PC を分離できる。それ以外にも、

- 録画データの管理が容易
- ネットワークに送られるデータをそのまま記録するので、再生時にはそのまま送り出せば良い
- 画像を用途に応じて縮小したり、複数カメラの画像を並べて 1 枚の画像にして録画することが可能

といった利点がある。逆に欠点としては、録画時の負荷が集中してしまう点が問題であるが、専用の高速・大容量 PC を割り当てることで、対応できる。

単一バイナリが複数の機能をはたすことにより、メンテナンスの負担を少なくしている。また、起動の際にコマンド引数として xml ファイルを http で指定してどの機能を使うのかを指定している。これらのファイルは起動時に一度読めばよく、また web サーバ上に集中管理できるため、メンテナンス性も良い。

図 4 は過去の記録・再生を行わない通常 TV 会議システムとして使用する際のシステム構成である。この場合はキャプチャ機能からネットワークを通して表示機能にデータが直接送られる。

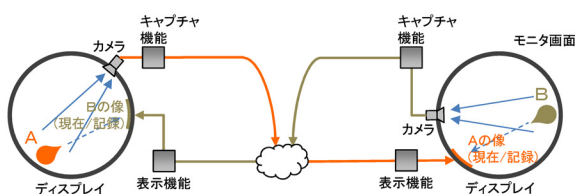


図 4 通常会話時のシステム構成

図 5 は録画・再生を行う時のシステム構成である。この時はキャプチャ機能の後に proxy 機能が挿入され、送られてきたデータを記憶装置に保存すると同時に表示機能に向けてデータを送る。また、再生時は記憶装置に保存されたデータを表示装置に送る。

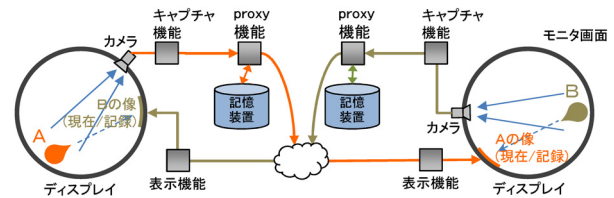


図 5 録画・再生時のシステム構成

図 6 は画像変換を行う際のシステム構成例である。キャプチャ機能からのデータは、1 つはそのまま表示機能に送られ、もう一つは proxy 機能に送られる。proxy 機能では受け取った画像を縮小してから表示機能に送る。例えばノート PC は、t-Room システムに比べて画面サイズが小さく CPU 性能も低い。ノート PC でサイズ縮小するのではなく他の高速な PC で動く proxy 機能が最適なサイズに縮小したデータを送ることでノート PC の負荷を減らすことが可能である。

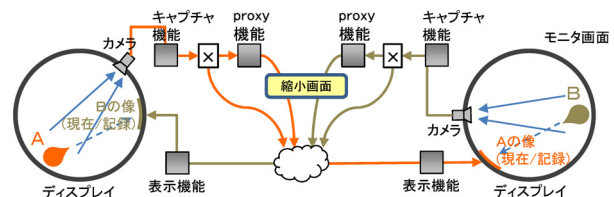


図 6 通常画面とノート PC (縮小) 同時使用時のシステム構成

proxy 機能はカメラ 1 台につき 1 つ設けていることも、数台まとめてもよいし、あるいは全 t-Room で 1 つでも良いが、負荷やネットワークの遅延を考慮し、各地点の t-Room 1 セットごとに 1 つ proxy 録画専用 PC で記録を行っている。

各 t-Room1 台につき、映像制御用のローカルコントローラを設置している。これは、どの過去を再生するのかといった選択機能や、映像の録画・再生・一時停止といった制御機能を受け持つ。複数台の t-Room を使って通信している様子を録画・再生するために、総合コントローラを用いる。総合コントローラはローカルコントローラに命令を送ることで複数台の t-Room をまとめて制御することが可能となっている。制御ソフトウェアの画面を図 7 に示す。ローカルコントローラは基本的にはどの PC 上でも動作するが、録画ファイルにアクセスする必要があるため録画している PC 上で動かしている。また、総合コントローラは ruby で実装されており、ローカルコントローラにアクセス

さえできればどの PC 上で動くようになっている。現在のところ、ノート PC 等のデバイスを用いて総合コントローラから制御を行っている。

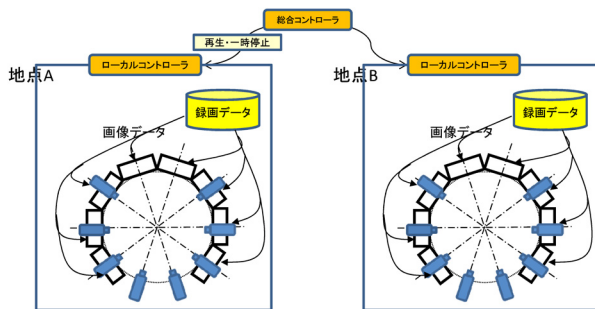


図 7 t-Room コントローラの配置

5.2 コヒーレントな再生と非コヒーレントな再生

本システムにより可能となったコヒーレントな再生方式の様子を図 8 に示す。図にあるように、コヒーレントな再生においては複数時点の過去の映像や遠隔地のリアルタイムの映像も全て同じように再生され、その見た目においては区別がつかない。再生された過去の情景に登場する人物も、その時に居た場所にいる状態で再生され、人物間の位置関係も保たれている。複数時点の過去を同時に再生する場合も (例えば過去の映像を見ている状態を録画した過去の映像の再生)、それぞれの空間的な位置関係や映像の大きさが異なることなく再生される。



図 8 コヒーレントな再生

このようにコヒーレントな再生は複数の過去を区別なく再生するが、一方でその区別をつけたい場合、例えばどの時点の過去を再生するかをブラウザして選択する等の用途には不向きである。そのため本システムでは、非コヒーレントな再生方法も実装している。非コヒーレントな過去再生では、過去映像を縮小してパネル状にし、その表示位置を変えることで複数の過去を分別可能な状態で表示する。図 9 にその一例を示す。水色の枠で囲まれた矩形領域がひとつの時点の過去映像となっている。過去映像のパネル

の大きさや位置は任意に変更可能となっており、保存されている過去映像一覧のブラウザや、過去映像の検索結果の表示など、その用途に応じて表示をすることを可能にしている。



図 9 非コヒーレントなパネル表示

5.3 ユーザインタフェース

現在、t-Room の基本操作はタブレット PC 等のウェブインタフェースを使用し、その画面に表示されたメニューをタッチして選択することで行う。t-Room の起動や終了、接続先、背景画像といった基本的な操作はすべてタブレット PC だけで指定可能である。タブレット PC は各地点の t-Room システム 1 セットに付き 1 台用意されており、起動前も起動後も、どこからでも制御可能となっている。

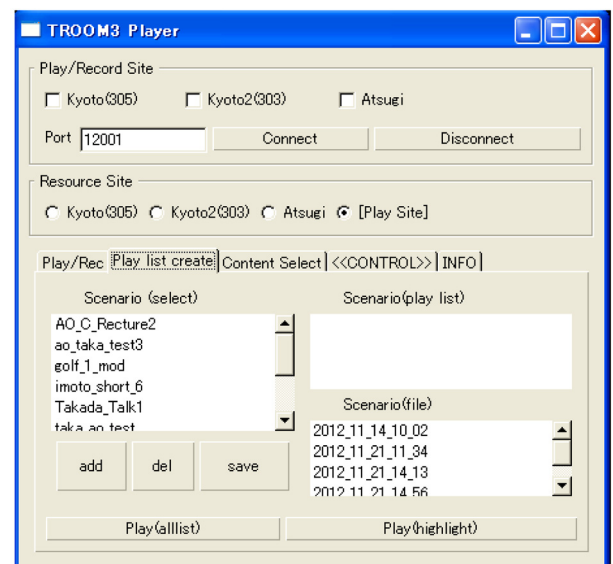


図 10 コントローラソフトウェアの画面

映像の記録や再生については、現在はまだプロトタイプの段階であり、このメニューには組み込んでおらず、PC からの操作となる。操作 PC の画面を図 10 に示す。この画面からどの地点の t-Room と接続するかを指定し、録画の指定、あるいは過去映像のリストからどれを再生するか

選択して再生開始・終了といった制御を行う。

6. 議論

6.1 録画機能の方式

過去情景の再生を行っている時の様子を更に録画する場合、次の2通りの方式が考えられる。

- (1) 録画時に再生している過去の映像と現在の様子を重ね合わせ、新たな録画ファイルを作成する
- (2) 録画時には現在の様子だけを録画しておき、再生時には録画時に再生していた過去映像と重ね合わせて再生する

1 の場合、再生している過去映像と現在のカメラ映像を重ね合わせた映像を作成するため、映像録画時の負荷は高くなるが、再生には一切余分な情報は必要がなく、負荷も軽い。しかし、映像は重ね合わせたものを保存していくため、何度も再生・録画を繰り返すと画質が劣化していく。一方、2 の方式では、過去映像の再生時に、その映像を録画していた時に再生していたより古い過去の映像を重ね合わせて再生する。この方式では再生時に多くの過去映像を重ね合わせることになり、また録画していた時に再生していた映像の情報も併せて管理しなければならない。ただし、映像を重ねて新たな過去映像を作成せず、録画したオリジナルを再生に用いるため、過去映像は劣化することはない。どちらの方式も一長一短であるが、現在、t-Room では重ね合わせの負荷が高くなるため 1 の方式を用いている。

6.2 過去映像の再利用

コヒーレントな映像記録・再生システムを用いると、記録された映像を用いつつ、それを再生しながらコメントをつけることができ、またそのコメントをつけた映像も録画し相手に送るといったことも可能になる (図 11)。

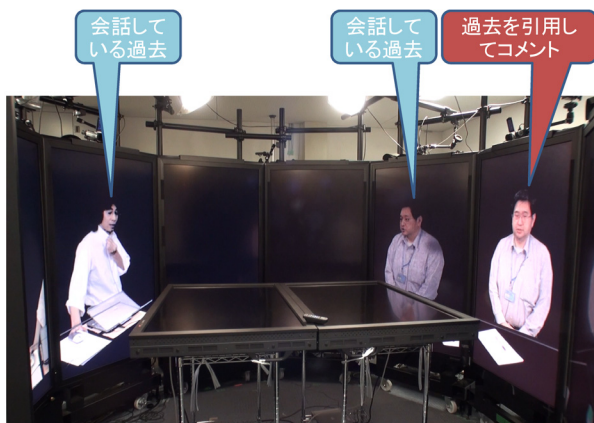


図 11 過去映像の引用例

時差のあるグローバルな仕事環境では、たとえ TV 会議システムを用いても相手に合わせて深夜に会議を行うなどしなければならないが、この録画/再生機能を使えば、自

分の好きなときに相手からの映像を見ることができ、また映像にコメントをつけて送り返すといった、いわゆる電子メールのような使い方が可能となる。これにより時差のある相手とのやりとりにおいても、自分の好きな時に相手からの映像を見たり、参照ビデオメールを作成して返信することが可能になる。また、過去の会議や会話を全部記録しておくことで、会話や作業の途中で困りごとが生じた際に、過去の情景を再生することで問題解決のヒントを得られる環境を作り出すことも可能となる。

TV 会議以外にも、スポーツレッスンや身体表現を用いたエンターテインメントなどの応用も考えられる。例えばゴルフレッスン等で遠隔地にいるプロに、生徒のスイングの映像を送り、プロが好きな時にフォームのチェックをした映像を作成したものを生徒がまた見るといった使い方も考えられる。他にも、遠隔地間の楽器演奏の際、現状ではネットワーク遅延や、演奏者が同時に集まることができないといった問題が生じることがある。コヒーレントな記録・再生を用いると、演奏者が好きなときに好きな場所で、過去の映像に合わせて演奏し、その映像をまた録画して重ねることで一つの映像作品を完成させるといったことも可能になる [15]。

7. まとめ

本稿では多地点 TV 会議システムにおけるコヒーレントな過去情景再生機能について述べた。この過去映像再生方式では過去映像と遠距離映像を同じように扱うことができ、過去の映像の途中に、コメント映像を挟んだりすることが可能であるため、従来の単純に再生するだけの過去情景の記録から過去を引用・再利用することが可能となる。また、コヒーレント再生のみでは複数過去の判別やリアルタイムな遠隔地との区別がつかなくなるため、コヒーレントでないパネル表示も同時に実装した。これらを使い分けることにより、今後は画像による身体動作認識を用いたユーザインタフェースを作成し、t-Room 内から過去映像の制御を行えるように実装し、また過去映像の検索機能、編集機能を充実させ、過去情景の再利用を進めるためのシステムの設計・開発を行っていく予定である。

参考文献

- [1] Keiji Hirata, Toshihiro Takada, Yasunori Harada, Shigemi Aoyagi, Yasuhiro Yamamoto and Kumiyo Nakakoji: "An Approach to Expressing and Sharing Value in Collaboratively Producing Creative Knowledge Artifacts," Proceedings of the 11th International Conference on Human-Computer Interaction, 2005.
- [2] Keiji Hirata, Yasunori Harada, Toshihiro Takada, Naomi Yamashita, Shigemi Aoyagi, Yoshinari Shirai, Katsuhiko Kaji, Junji Yamato and Kenji Nakazawa: "Basic Design of Video Communication System Enabling Users to Move Around in Shared Space," IEICE Transactions on Electronics, vol.E-92-C, No.11, pp.1387-1395, 2009.

- [3] Naomi Yamashita, Hideaki Kuzuoka, Keiji Hirata, Shigemi Aoyagi and Yoshinari Shirai: "Supporting fluid tabletop collaboration across distances," Proc. CHI'11, pp.2827-2836.
- [4] Robert S. Fish, Robert E. Kraut and Barbara L. Chal-fonte: "The VideoWindow system in informal commu-nication," Proceedings of the 1990 ACM conference on Computer-supported cooperative work, pp.1-11,1990.
- [5] John C. Tang, Scott L. Minneman: "VideoDraw: a video interface for collaborative drawing," Proceedings of the SIGCHI conference on Human factors in computing sys-tems, pp.313-320,1990.
- [6] 原田 康徳: "同室感通信", インタラクティブシステムとソフツウェア VI, pp.53-60,1998, 近代科学社.
- [7] Hideaki Kuzuoka, Jun Yamashita, Keiichi Yamazaki and Akiko Yamazaki: "Agora: a remote collaboration system that enables mutual monitoring, " CHI '99 extended ab-stracts on Human factors in computing systems, pp.190-191,1999.
- [8] Osamu Morikawa and Takanori Maesako: "HyperMir-ror: toward pleasant-to-use video mediated communi-cation system," Proceedings of the 1998 ACM conference on Computer supported cooperative work, pp.467-476, 1998.
- [9] Steve Harrison, Scott Minneman, Bob Stults and Karon Weber: "Video: a design medium," ACM SIGCHI Bul-letin, vol.21, no.2, pp.62-67, 1989.
- [10] Sara A. Bly, Steve R. Harrison and Susan Irwin: "Media spaces: bringing people together in a video, audio, and computing environment," Communications of the ACM, vol.36, no.1, pp.28-46, 1993.
- [11] Jeremy R. Cooperstock, Sidney S. Fels, William Buxton and Kenneth C. Smith: "Reactive environments," Com-munications of the ACM, vol.40, no.9, pp.65-73, 1997.
- [12] Patrick Chiu, Ashutosh Kapuskar, Lynn Wilcox and Sarah Reitmeier: "Meeting capture in a media enriched conference room," Proceedings of CoBuild '99, pp.79-88, 1999.
- [13] Stefan Agamanolis, Alex Westner and V. Michael Bove, Jr.: "Reflection of presence: toward more natural and responsive telecollaboration," Proceedings of SPIE Mul-timedia Networks, vol. 3228A, 1997.
- [14] Toshihiro Takada and Yasunori Harada: "Citation-capable video messages: overcoming the time differences without losing interactivity," i3 Annual Conference Pro-ceedings, pp.13-15, 2000.
- [15] 梶 克彦: "いつでもだれとでもセッションできる、時空間を超えるビデオシステム t-Room," DTM マガジン, 2010年3月号, pp.88-89.