

Gigabit Ethernet カメラを利用した 超高精細全方位映像システム

大葛 広和 佐藤洋介 米田裕也 橋本 浩二 青木慎一郎 柴田 義孝
岩手県立大学ソフトウェア情報学部 岩手県立大学社会福祉学部
株式会社アイソニック

本稿において高解像度全方位映像を用いた多人数参加型の遠隔テレビ会議システムについて述べる。近年、遠隔地から会議へ出席する目的でTV会議システムを導入する例が増えている。しかしながら従来のTV会議システムに利用されるカメラは撮影範囲が狭く、1度に広い範囲の撮影や注目する視野全体を提供することは困難であった。筆者らは過去の研究において、これらの問題点を解決するために全方位カメラを導入し、360度の視野を有するTV会議システムを構築した。しかしながら、過去に開発したシステムでは表示解像度が限られているため、発話者の特定や発言者の表情をとらえにくい問題があった。そこで本稿では高解像度 Gigabit Ethernet カメラを利用して細部映像まで分析可能な高解像度全方位映像を有する多人数参加型の遠隔TV会議システム(超高精細全方位映像システム)を構築し、本システムの実現方法とプロトタイプシステムについて述べる。

High-Resolution Omni-directional Video System with Gigabit Ethernet Camera

HIROKAZU OHKUZU, YOSUKE SATO, YUYA MAITA, KOJI HASHIMOTO,
SHIN-ICHIRO AOKI YOSHITAKA SHIBATA,
Faculty of Software and Information Science, Iwate Prefectural University
Faculty of Social Welfare, Iwate Prefectural University
AiSONIC Co Ltd.

In this paper, we introduce a new teleconference system by using high-resolution Omni-directional video. Recently, there has been increased number of introduction of teleconference systems in the office to join to a conference from remote site. However, since those teleconference systems usually use conventional cameras which only capture the objective images in one direction with narrow area, more wide area such as 360 degree of surround direction cannot be captured. In the previous research, we developed the teleconference system which could capture a wide area such as 360 degrees with omni-directional camera for the purpose of solving these problems. Because omni-directional video was low resolution, the teleconference system which we developed in the past was not suitable for the detailed analysis of the human face. Therefore we developed the teleconference system which could capture high-resolution omni-directional video suitable for a detailed analysis using a high-resolution Gigabit Ethernet camera. In this paper, we describe an architecture and implementation method of our system and prototype.

1. はじめに

近年、通信ネットワークの高速化とマルチメディア処理技術の発達に伴い、IPベースのネットワーク上で高品質の映像や音声を用いて、遠隔地間をリアルタイムかつ双方向でコミュニケーションを可能とする遠隔講義システムや遠隔TV会議システムが研究され、その一部は実用化されて来ている。また、企業の国際化が進む中で、遠隔地への会議出席は、多大な交通費や宿泊費、そして移動時間を要し、さらに不在時に業務が停滞するなど、仕事効率面において多くの問題を発生させる。これらのことから、遠隔TV会議システムを導入する例が増えており、多数企業において日常的に遠隔TV会議システムが使用されている。現在、商用の遠隔TV会議システムは、1拠点多人数参加型が主流となっている。グローバルマーケットにおいてTV会議システ

ム販売台数1位のPOLYCOM社の製品は、10~15名以下の参加人数対象としている。他メーカーも同様の参加人数を対象とした製品を提供している。このようなシステムのほとんどは従来の単一方向型カメラを利用しているため撮影範囲が狭く、一度に広い範囲の映像の撮影が困難であるため、多人数参加型の遠隔TV会議において各参加者の映像が小さくなってしまい、「発話の特定」や「発言者の表情をとらえにくい」などの問題があった。そこで筆者らは、これまで、一度に360度撮影可能な全方位カメラを導入し、取得した全方位環状映像をパノラマ映像に展開し、これをマルチメディアストリームとしてRTPパケットでリアルタイムに送信し[2]、受信側でパノラマ映像を表示したり、その中の注目する領域を拡大したり、ゆがみを補正する機能を有する全方位映像通信ミドルウェア[1]を開発してMidField System[2]との組み合わせによる全方位

映像利用の遠隔テレビ会議システムを実現してきた。しかしながら、過去に開発したシステム[3]では全方位映像の表示解像度が限られているため、映像の詳細分析を行うために PTZ (Pan/Tilt/Zoom)カメラやDVカメラを組み合わせる必要があり、システムが複雑化していた。

本研究では、高解像度 Gigabit Ethernet カメラ[4]と全方位映像通信ミドルウェア[1]を利用して超高解像度全方位映像を有する遠隔 TV 会議システム(超高精細全方位映像システム)を構築する。超高精細全方位映像は 360 度視野を有するだけではなく、映像の詳細分析利用も可能であるため多様な利用者環境やアプリケーションでの活用が期待できる。本稿では、提案するシステムの構成と要素技術、利用分野の提案、前述の機能を実現するプロトタイプシステムについて述べる。

2. システム概要

本システムの構成は図1で示すように、高速ネットワーク上に存在する全方位映像処理用サーバ PC とクライアント PC で構成されている。サーバ PC は全方位環状映像のパノラマ映像展開処理、画像補正、画像符号化機能を備えている。

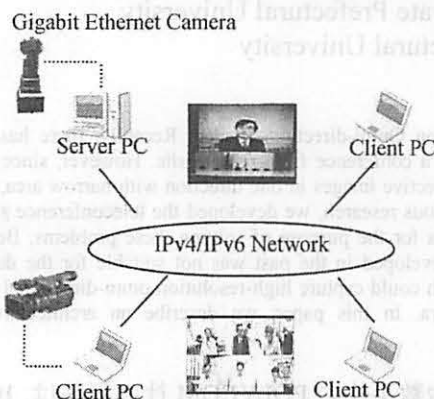


図1 システム構成



図2 小型モニターと全方位カメラ(サーバ装置)

全方位映像は PAL レンズを装着した Gigabit Ethernetカメラからギガビットネットワークを利用してサーバ PC に取り込まれる。取り込まれた全方位映像はパノラマ展開処理を施し

MidField System[2]を用いて DV, WMV の圧縮符号化方式でクライアント PC へ配信される。クライアントではサーバ側から配信された映像の表示と DV カメラを使用してサーバへ単方位映像(DV)配信を行う。本システムではクライアントから配信された単方位映像表示用に小型モニター(図2)を複数枚用いている。

2.1. 全方位映像

全方位カメラはいくつかのメーカーより発売されているが、本研究での全方位映像の撮影は高解像度Gigabit EthernetカメラにPALレンズ(図3)を装着し、ギガビットネットワークでPCに接続することで全方位映像を取り込む。取り込まれる全方位映像を環状画像と言い2048×2048ピクセル、15fpsで周囲360度と垂直方向に50度(水平を基準に-20度から+30度)の視界を有するドーナツ状の映像である。このドーナツ状の環状画像をサーバPC内の全方位映像ミドルウェアによる展開処理を施しユーザにとって容易に理解可能なパノラマ映像を配信する。

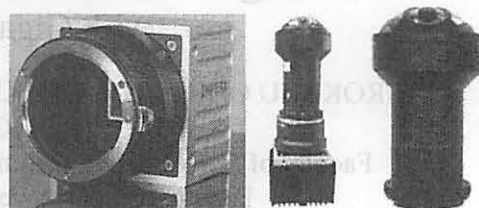


図3 全方位Gigabit EthernetカメラとPALレンズ

2.2. 画像処理

PAL レンズを用いて撮影される全方位映像はドーナツ状の環状映像(図4環状画像)であり、そのままの形状では視覚的に理解しにくいので次に示すような360度のパノラマ画像(図4パノラマ画像)、部分拡大(デジタルズーム)画像、180度2分割画像への展開処理が可能である。また、本システムで利用する全方位カメラは撮影環境に応じた適切なカラーバランスが自動的に設定されないため、ユーザ要求に応じて各画素値と輝度値のパラメータを設定可能としている。



図4 全方位展開処理

2.3. Gigabit Ethernet カメラ用インターフェース

筆者らは Gigabit Ethernet カメラ(図2)から得られる映像データを MidField System[2]で利用するためのインターフェースを構築した。このインターフェースは Dynamic Link Library(DirectShow フィルタ)として構築されて、Windows OS 上で多様なアプリケーションに組み込んで利用することが可能である。このインターフェースは Gigabit Ethernet を利用してカメラと接続し、カメラ設定、出力フォーマットの取得、映像データの RGB 変換を行う。本システムで使用するカメラの撮影素子(CCD)は光の強度に関する濃淡情報しか得ることができない。そこでカラー化のために CCD 素子にカラーフィルタを被せ、個々の CCD 素子に色を認識させている。カメラからデータを取得した後にソフトウェアで信号処理を行うことで個々の画素をカラー化させる必要がある。カラーフィルタのパターンは図5の様な Bayer パターンである。色再現法はいくつか存在するが本システムでは線形補間法を使用している。

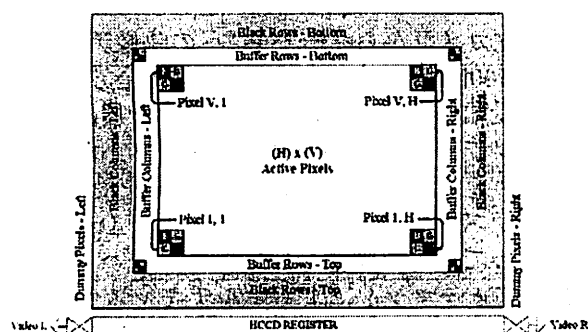


図5 カラーフィルタパターン

3. システムアーキテクチャ

本システムは図6に示すように5層(アプリケーション層、データトランスフォーム層、MidField層、GigaEインターフェース層)から構成される。

- ・アプリケーション層はTV会議のアプリケーションでありユーザインターフェースの提供と映像の表示を実現している。

- ・データトランスフォーム層では全方位環状画像のパノラマ、部分拡大、補正画像への展開処理、及び各画素値と輝度の調整、映像の分割処理を行う。

- ・MidField層では後述するように、映像転送のためのミドルウェアでありネットワークを通じてメディアパケットの転送を行う。

- ・GigaE インターフェース層では Gigabit Ethernet カメラの接続とカメラ設定、MidField System への映像提供を実現している。

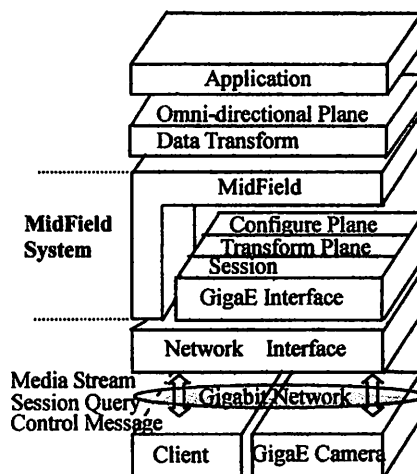


図6 アーキテクチャ

4. MidField System

MidField System[2]は、通信端末の処理能力や利用可能なネットワーク帯域に応じて、適切なフォーマットによる通信を実現する柔軟性を持ったマルチメディア相互通信ミドルウェアである。多様なユーザ環境において、ユーザの QoS 要求を考慮しトランスコーディング機能を有する通信サービスの提供を可能とする分散マルチメディアシステムである。

5. 全方位映像ミドルウェア

全方位映像ミドルウェア[1]とはデータトランスフォーム層の全方位環状画像のパノラマ映像展開処理を実現している。筆者らが開発した全方位映像ミドルウェアは多様な入力解像度(DV, HDV 以外にも 2048x2048 ピクセル, 1600x1200 ピクセル)に対応している。また出力解像度も本システムで使用する全方位レンズの仕様に合わせて 2880x460 ピクセル, 2000x350 ピクセル, 1200x200 ピクセルを用意した。

5.1. 全方位映像展開プロセス

全方位レンズを装着した Gigabit Ethernet カメラから取得した全方位映像の展開プロセスは図7に示すような形で行われる。

1. 全方位映像展開用座標テーブルを作成する。
2. パノラマ画像・部分拡大画像・補正画像へ

- の展開を行う。座標変換テーブルを参照し環状画像から必要なピクセルを抽出する。
- 抽出したピクセルの近隣4ピクセルを参照し線形補間処理を適用する。
 - 2と3を繰り返すことによって全方位映像が作成される。

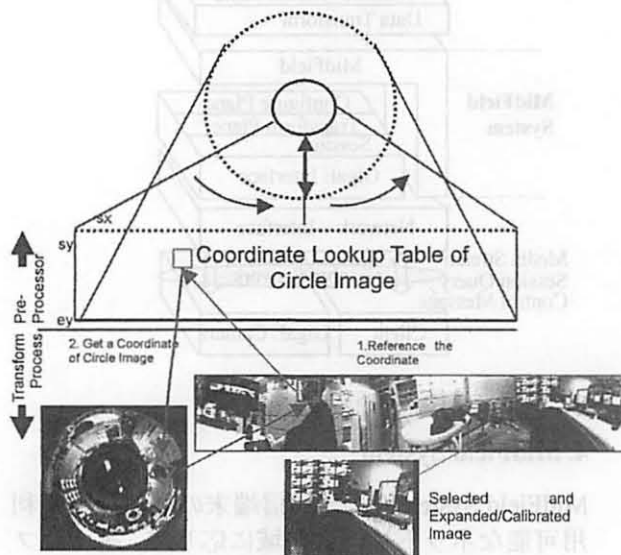


図7 全方位画像の展開処理

6. プロトタイプシステムと性能評価

筆者らはプロトタイプシステムとして遠隔テレビ会議システムを構築し実験を行った。使用プログラミング言語は C++(Microsoft Visual C++2005)を使用しメディア処理と管理には DirectX9.0b(DirectShow)を用いて開発した。

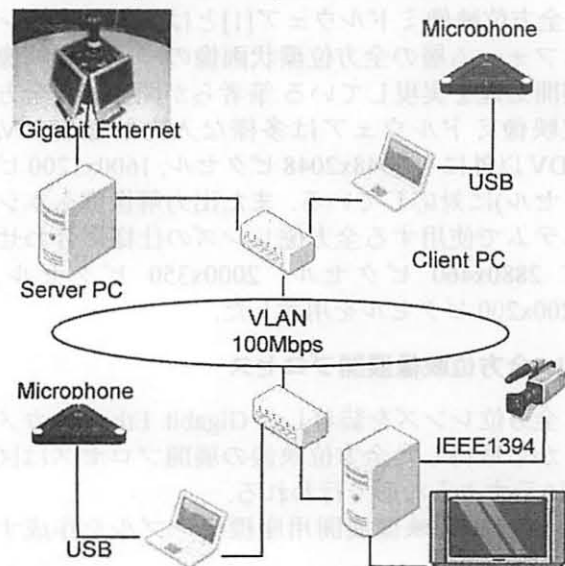


図8 プロトタイプシステム

プロトタイプシステムはサーバ PC と全方位レンズを装着した Gigabit Ethernet カメラ、4 台の小型モニター、無指向性マイク、クライアント PC、DV カメラ、大型プラズマディスプレイ、Gigabit Ethernet スwitchングハブと VLAN ネットワークで構成される。(図 8)

6.1. 性能評価

本システムの性能を示す指標として使用ネットワーク帯域、CPU 使用率、映像の解像度とフレームレートについて評価を行った。

サーバ PC とクライアント PC の性能は表 1 に示している。サーバ PC は Gigabit Ethernet カメラを直接、ギガビット対応の LAN ケーブルで接続するためネットワークカード(NIC)を 2 枚有する。

Sepecification	Server	Client
CPU	Xeon quadcore 3.2GHz x 2CPU	Intel® Core™2
Main Memory	8GB	3.25GB
NIC	1Gbps	1Gbps
NIC	1Gbps	

表1 PC 性能

全方位カメラからサーバ PC への映像データ量は平均 500Mbps である。サーバ PC で全方位カメラから受信した全方位映像はパノラマ展開されて WMV フォーマットでマルチキャスト配信する。クライアント PC では全方位映像を受信して大型プラズマディスプレイへの表示と DV カメラによって単方位映像を撮影してサーバ PC へ WMV フォーマットによるマルチキャスト配信を行う。プロトタイプシステムでは、この方法によって双方向映像通信を実現した。実験におけるシステムの性能を表 2 に示している。

Items	Server	Client
Video Format	Omni-directional	DV
Pixel Resolution	2000x350	720x480
Codec	WMV	WMV
Req. BW	4Mbps	4Mbps
Frame Rate	11fps	29fps
Ave. CPU load	24%	58%

表2 システム性能

実験ではサーバ PC での全方位映像出力解像度とフレームレートの関係についても検証した。表 3 に実験で使用した出力解像度とフレームレートの関係を示している。

出力解像度	画像補正(有)	画像補正(無)
2880x460	6.9fps	7.0fps
2000x350	11fps	11.2fps
1200x200	13.5fps	13.7fps

表3 解像度とフレームレート

全方位映像は高解像度になるとフレームレートが減少するため利用用途に応じた出力解像度を選択する必要がある。テレビ会議システムでの全方位映像利用はスムーズで精細な映像が必要である。そこでプロトタイプシステムでは出力解像度 2000x350 ピクセルを選択している。

6.2. DV カメラ全方位映像と比較

DV カメラと高解像度 Gigabit Ethernet カメラを利用して全方位映像を作成した場合の違いについて比較した。DV カメラを利用して全方位映像を生成した場合、DV(720x480)では1フレームあたりの処理時間は 25ms でフレームレートは 30fps で展開処理が可能であり、HDV(1440x480)では1フレームあたりの処理時間は 47ms でフレームレートは 13fps で展開処理が可能であった。一方、高解像度 Gigabit Ethernet カメラを利用した場合、高解像度映像(2048x2048)を 1 フレームあたりの処理時間 29ms でフレームレートは 11fps で全方位映像(2000x350)に展開処理が可能である。また、解像度 2880x460 ピクセルと 1200x200 ピクセルへの展開処理時間は 1 フレームあたり、それぞれ 50ms と 14ms で展開処理が可能であった。画質は DV カメラを用いた場合の解像度が 800x200 ピクセルであるが、Gigabit Ethernet カメラを利用した場合、最大解像度は 2880x460 ピクセルにまで改善された。図9に DV カメラによる全方位映像と高解像度 Gigabit Ethernet カメラによる全方位映像の比較を示している。

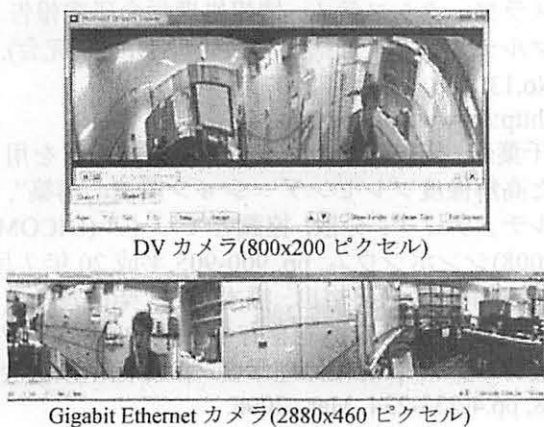


図9 従来システムと比較

また、DV カメラと PC を接続するために必要な IEEE1394 ケーブルの最大距離は 4.5 メートルであるが Ethernet ケーブルは接続距離制限を考慮する必要がないため、会議室内の自由なレイアウトが可能である。近年、IEEE802.11n 対応の最大 300Mbps 以上の速度を実現する無線ルータが市場に出回っており、将来、無線環境を利用して Gigabit Ethernet カメラと PC の接続が可能となる見込みである。高解像度 Gigabit Ethernet カメラを利用することで解像度と接続インターフェースの点で多くの利点が存在する。

6.3. 音声処理機能

プロトタイプシステムは、話者の声を集音するためにエコーキャンセラーを搭載した無指向性型マイクスピーカー(図 10)を利用する。無指向性型マイクスピーカーは話者の人数や方向に応じて可動式マイクアームを動かすことによって集音範囲や指向性を変更することが可能である。PC に取り込んだ音声データは MidField System[2]を利用して音声ストリームとして配信される。



図10 無指向性型マイクスピーカー

7. 遠隔メンタルヘルスケア教育

次に超高精細全方位映像システムを遠隔メンタルヘルスケア教育に活用した事例を示す。多人数参加型の遠隔セミナーや遠隔教育を行う際、参加者の表情や動作を効率良く撮影する方法が必要である。そこで超高精細全方位映像システムを遠隔メンタルヘルスケア教育に導入して遠隔地に居る精神科医師の講師とセミナー参加者が対話を行った。セミナー参加者は 7 名であり円卓テーブルに座り、円卓中央のモニターに映し出される講師の映像を視聴する。(図 11)

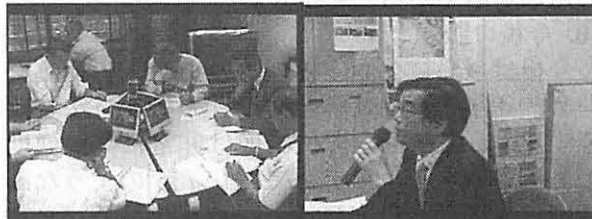


図11 遠隔メンタルヘルスケア教育



図 12 講師側視聴イメージ



図 13 セミナー参加者側視聴イメージ

図 12 と図 13 に講師とセミナー参加者が実際に視聴した画面イメージを示している。超高精細全方位映像によって講師がセミナー参加者の詳細な表情や動作を分析可能であることが確認できた。

9. まとめ

本稿では、超高解像度 Gigabit Ethernet カメラ、全方位映像ミドルウェア、MidField System を利用した超高精細全方位映像システムの概要と映像処理、性能評価、プロトタイプシステムによる会議システムの構築と評価結果について述べた。プロトタイプシステムでは遠隔メンタルヘルスケア教育セミナーで実際にシステムを利用して評価を行った。その結果、これまでのシステムと比較し、システム機能や解像度の点で大幅に改善でき、またこれまで不可能であった遠隔メンタルヘルスケアや医療ヘルスケアへの応用が可能であることがわかった。

10. 今後の課題

今後はさらにプロトタイプシステムの構築を進め、全方位映像表示と発話者の特定方法を改善する。超高精細全方位映像は非常に高解像度であるためタイルドディスプレイ[5]のような高精細で大画面かつ高臨場感を持つディスプレイ環境の利用を検討している。

タイルドディスプレイとは図 1 に示すような複

数の液晶ディスプレイを組み合わせたディスプレイ環境を指し、専用のグラフィックハードウェアを利用せずにソフトウェアを利用した仮想的な高解像度ディスプレイ環境を構築できる。(図 14)

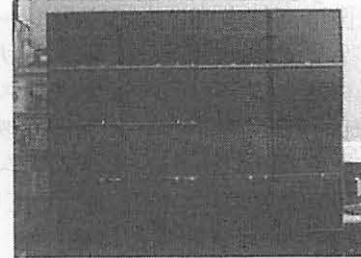


図 14 タイルドディスプレイ

また、発話者の特定方法について無指向性マイクを用いた場合、全ての方向から均等に集音するため全方位映像中の発話者の位置特定には不向きであった。そこで複数の有指向性マイクと音声方向決定アルゴリズム[6]を用いることで話者の方向を検出し、現在の話者方向の抽出や表示、及び話者方向の音声のみの配信を実現する予定である。さらに、上記に述べた機能の追加や超高精細全方位映像システムの新しい応用分野についても検討する予定である。

References

- [1] 米田裕也, 橋本浩二, 柴田義孝: 高精細全方位映像の利用と通信のためのミドルウェアの開発, 情報処理学会第 68 回全国大会, pp.581-582, (2006)
- [2] 橋本浩二, 柴田義孝: 利用者環境を考慮した相互通信のためのミドルウェア, 情報処理学会論文誌, Vol. 46, No. 2, pp.403-417, (2005)
- [3] 佐藤洋介, 米田裕也, 橋本浩二, 柴田義孝: 全方位カメラと制御カメラを利用した遠隔カメラワークシステム, 情報処理学会研究報告マルチメディア通信と分散処理(DPS 研究会), No.130, pp. 429-434 2007 年 3 月
- [4] <http://www.imperx.com/>
- [5] 千葉豪, 柴田義孝, “マルチディスプレイを用いた高解像度プレゼンテーション環境の構築”, マルチメディア, 分散, 協調とモバイル(DICOM O 2008)シンポジウム, pp. 900-905, 平成 20 年 7 月
- [6] 斎藤純一, 米田裕也, 橋本浩二, 柴田義孝, “発話者の位置特定機能を有する全方位映像 TV 会議システム”, 情報処理学会 第 68 回全国大会, 5T-8, pp.4-233-234, Mar., 2006.