

音声映像情報の多点統合による情景観測システムの研究 (041306003)
Research on Scene Viewing System
Based on Multi-Point Integration of Audio and Visual Information

研究代表者

間瀬健二 名古屋大学情報連携基盤センター
Kenji Mase Information Technology Center, Nagoya University

研究分担者

武田一哉† 藤井俊彰†† 平野靖†††
Kazuya Takeda† Toshiaki Fujii†† Yasushi Hirano†††
†名古屋大学大学院情報科学研究科
††名古屋大学大学院工学研究科
†††名古屋大学情報連携基盤センター
† Graduate School of Information Science, Nagoya University
†† Graduate School of Engineering, Nagoya University
††† Information Technology Center, Nagoya University

研究期間 平成 16 年度～平成 18 年度

本研究開発の概要

舞台演奏・演技、車内運転状況などの実時空間情報を任意の位置で視聴・鑑賞・分析可能とする、多点統合可能なマルチメディア情景観測システムを研究開発する。具体的には、高品位音声および高解像度映像を、高精度同期して多点記録するシステムを開発した。また、本システムを用いて実時空間情報を実際に収集し、光線空間法に基づく実写ベースの自由視点画像再生技術、頭部音響伝達特性を利用する空間音場再生技術の統合および多点信号の要約技術等開発のための基礎データを収集分析した。これら、多点観測のための観測モジュールおよびその多点同期技術の開発と、100チャンネル規模の音声映像の多点統合および自由視聴点音像生成技術をもとに、音声と映像を統合した自由視聴点コンテンツの生成を可能とした。

Abstract

The aim of our research project is to develop a multimedia scene viewing system that can integrate multi-point audio- and visual-information for a free-view point audio-visual content creation. We focus on a real-world space-time information recording and communication system, which can view, appreciate, and analyze the real-world information such as stage-performances, plays, and driving situations from arbitrary viewing point. We have developed a 100 point audio-visual recording system and succeeded in create a free view point content of audio-visual integrated scene.

1. まえがき

ネットワークのブロードバンド化により、講義や競技スポーツの放送中継をユーザの好みのカメラ映像で鑑賞することが、可能となりつつある。しかし、視聴者にとっては、いまだコンテンツ提供者の視点（すなわちカメラやマイクロフォンの位置）を通してのみ、その情景を鑑賞する形態しか実現されていない。自由な視点から、視聴者にとって価値のある情報に容易に素早くたどりつき、編集・要約・閲覧しながら他者と共有できる、次世代情報通信技術が望まれている。

本研究開発課題では、当グループが有する自由視点映像合成などの多点観測データの統合・合成技術を基盤として、高品位音声と高解像度映像を含む情景を、高精度に多点同期して観測可能なマルチメディア情景観測システムを開発した。また、多点音声信号からの自由聴点音像の生成技術を新たに開発した。さらに、このマルチメディア情景観測システムを使って、実際に音楽演奏、ピアノなどの実世界情景データを収集し、自由視聴点から音声と映像が同時に変化する臨場感あふれる音声映像コンテンツの生成に成功した。

2. 研究内容および成果

2.1 マルチメディア情景観測システムの開発

自由視聴点での映像コンテンツの生成のためには、音声と映像からなるマルチメディア情景を、高精度に多点同期してセンシングして記録できる情景観測システムが必要である。そこで、本研究課題では、次の性能を有する情景観測システム[1]を開発した。

【センシング基本性能】

- ・ 超高解像度ビデオカメラインタフェース（映像）
解像度：1392(H)×1040(V)、カラー（Bayer 方式）
フレームレート：29.4118 fps
- ・ 複数チャンネル A/D 変換インタフェース（音声ほか）
サンプリングレート：最大 96kHz、2～4 ch

【同期性能】

- ・ GPS フレーム同期制御ユニット
各ノード間同期精度：1μ秒（有線）
10m 秒（無線）

高精度でデータ収録を実現するためのシステムの基本設計と実装を行った。図 1 に 100 台のノードを配置した例を示す。カメラ・マイクの配置は被写体に応じて柔軟に配置が可能な構成となっており、直線配列、円弧配列、2

次元配列などを検討した。

また、フレーム同期制御ユニットから発生された信号を、大規模数の各モジュールに遅延なく分配する仕組みと同期信号結線のトポロジーについて検討した。クラスタ間の同期をとるために GPS の基準信号を利用し、遠隔地にあるフレーム同期制御ユニット同士の内部クロックを同期させることで、有線接続のみならず、GPS 基準信号による遠隔のクラスタ間での同期が可能となった。



図 1 情景観測システム
(100 ノードの円弧型配置例)

2.2 自由聴点音像生成アルゴリズムの提案

画像の自由視点合成で用いられる光線空間法における光線に関するサンプリング定理であるブレンオプティク理論を援用して、音空間に対してマイクロホンアレイ (MA) のサンプリング間隔に関する定理を提案した。また、音波形の光線空間表現とサブバンド信号処理をもちいて、音源定位を行うことなく任意聴点の生成を可能とするアルゴリズムを提案した[2]。

まず、受音における音場を光線空間の形式で表現するため、複数マイクロホンによるアレイビームフォーマを構成した。その際それぞれの MA が異なる方向にビームを向けることにより、音像を広い周波数帯域に渡り生成できることを示した。それぞれの周波数帯域の信号は、特定の周波数帯域に最適に設計されたマイク間隔をもつ MA によって記録され、それによって音場の光線空間が生成される。ここで、MA 間の距離は完全な再構成のための標準化定理を満たすように設置する必要がある。理論的には、最大許容変位を音像中の 1 標本点にあたるように設定することが望ましい。

2.3 多点空間信号の伝送・処理方式の検討

複数カメラによるセンサーネットワークで取得した多眼画像の分散情報源符号化法を検討した。ノード間で統計的相関がある情報源について、従来はノード間での符号化をする必要があったが、本検討により、アダプティブなモジュールを使えば、情報源の位置で符号化をしても、符号化効率がとくに落ちないことをシミュレーションによって示した。これを発展させると、100 ノードの画像を使って自由視点映像を構成する際に、各ノードでの符号化をしてサーバにデータを集めた後に復号化する事が可能となる[3]。

2.4 自由視聴点コンテンツの作成

開発した情景観測システムを用いて音楽演奏シーンとピアノ演奏を収録し、自由視聴点から鑑賞できるコンテンツを生成した。

音楽演奏シーンの収録では、図 2 に示すように、音源である 2 名の演奏者を囲うようにカメラを 16 台配置し、また、48 本のマイクロホンによって構成されたマイクロホンアレイを正面の床に設置した。自由視点映像は、離散的なカメラ映像を元に光線空間法によって生成する。一方、自由聴点を実現するために、初めに、同期収録された 48

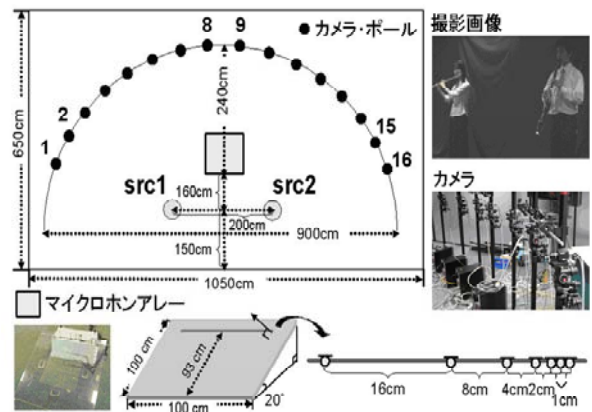


図 2 楽器演奏コンテンツの収録

チャンネルの演奏信号に対して周波数領域独立成分分析にもとづいた音源分離手法を適用することで、個々の楽器演奏信号を抽出した。個々の演奏信号の抽出にあたっては、歪を生じさせないという観点にたち、分離フィルタの設計を行った。続いて、分離した演奏信号とユーザが選択した受聴位置と各音源との間に対応する頭部伝達関とを畳み込み再合成を行った。

この手法により再合成を行うことで、ユーザが自らの視聴位置を自由に設定することが可能なシステムを実現した。生成された演奏映像は、視点とともに聴点の移動を実感できる臨場感の高さを確認できるコンテンツとなった。

3. むすび

本研究では、多点での連続映像・音声を高精度に記録する装置の開発と、従来の自由視点合成手法に対応する自由聴点合成手法を開発することにより、ユーザ中心の情報閲覧手段となる自由視聴点閲覧技術の進展に貢献した。とくに、実際に、自由な視聴点で音声と映像を同期させて鑑賞できるコンテンツを作成したことは、その効果を他の研究者やコンテンツ制作者が体感することができるようになった点で、波及効果は大である。

本成果の映像・音声を統合して自由視聴点を呈示する技術をもとに、MPEG では自由視聴点 TV の標準化を対象とすることが議論されている。

【誌上発表リスト】

- [1] Toshiaki Fujii, Kensaku Mori, Kazuya Takeda, Kenji Mase, Masayuki Tanimoto, Yasuhito Suenaga, "Multipoint Measuring System for Video and Sound -100-camera and microphone system-", IEEE 2006 International Conference on Multimedia & Expo (ICME2006), pp.437-440, (2006 年 7 月 11 日).
- [2] Mehrdad Panahpour Tehrani, Yasushi Hirano, Toshiaki Fujii, Shoji Kajita, Kazuya Takeda and Kenji Mase, "Arbitrary Listening-point Generation Using Sub-band Representation of Sound Wave Ray-space," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2006), Vol.5, pp.329-332, (2006 年 5 月 14~19 日).
- [3] Mehrdad Panahpour Tehrani, Toshiaki Fujii, Masayuki Tanimoto, "The Adaptive Distributed Source Coding of Multi-view Images in Camera Sensor Networks," IEICE Trans. on Fundamentals of Electronics, Communication and Computer Sciences, Vol.E88-A, No.10, pp.2835-2843 (2005 年 10 月).