

研究速報

多視点・多聴点データ取得システムを用いた自由視聴点映像生成

福嶋 慶繁<sup>†a)</sup>                      丹羽 健太<sup>††</sup>  
圓道 知博<sup>†</sup>(正員)                  藤井 俊彰<sup>†</sup>(正員)  
谷本 正幸<sup>†</sup>(正員:フェロー)      西野 隆典<sup>††</sup>(正員)  
武田 一哉<sup>††</sup>(正員)

Free Viewpoint and Listening-Point Video Generation by Using Multi Viewpoint and Multi Listening-Point Data Capturing System

Norishige FUKUSHIMA<sup>†a)</sup>, Kenta NIWA<sup>††</sup>, *Nonmembers*,  
Tomohiro YENDO<sup>†</sup>, Toshiaki FUJII<sup>†</sup>, *Members*,  
Masayuki TANIMOTO<sup>†</sup>, *Fellow*, Takanori NISHINO<sup>††</sup>, and  
Kazuya TAKEDA<sup>††</sup>, *Members*

<sup>†</sup> 名古屋大学大学院工学研究科, 名古屋市  
Graduate School of Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, 464-8603 Japan  
<sup>††</sup> 名古屋大学大学院情報科学研究科, 名古屋市  
Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, 464-8603 Japan  
a) E-mail: fukushima@tanimoto.nuee.nagoya-u.ac.jp

あらまし 本論文は、三次元の音声・映像を統合した新たなメディアを提案する。まず、多数のカメラ、マイクロホンと並べたカメラアレイ、マイクロホンアレイで、多視点・多聴点データを撮影した。次にそのデータより自由視点映像、自由聴点音声を生成し、自由視聴点映像の生成に成功した。

キーワード カメラアレイ、マイクロホンアレイ、自由視聴点

1. まえがき

近年、多数のカメラから三次元情報を蓄積し、その再構成を行う研究が盛んに行われている [1]。またそのデータから任意の視点を作り出すことができる自由視点画像生成の研究も多数行われている。

しかし、その三次元シーンの情報を取得・再現するとき、音声データまでを再現したものは少ない。本論文では、多数のカメラとマイクロホンからなる、カメラアレイとマイクロホンアレイを用いて、三次元情報を取得し、そのデータからユーザが任意の視点・聴点を選ぶことができる、自由視聴点映像生成を行う。

2. システム構成

本システムは、制御用のサーバと、画像、音声の取得用ノードから構成される。各ノードは1台のカメラ(1392 × 1040 の解像度, 30 fps)と4台のマイクロホンを接続可能である。本システムの主な特徴として、全映像・音声データ間の同期が1  $\mu$ s 以内でとれる点に

ある。もう一点として、音声・映像ともに非圧縮データが取得可能な点である。前者は、すべてのノードを同期線で直列に結ぶことで達成される。また、後者はシステム構成として、各カメラごとに1台のノードを割り当てることで取得可能としている。両性能ともに、信号処理を研究上で、圧縮された汎用的なものよりも有用なデータとなる。

これらのシステムにより、図1中にあるカメラアレイ、マイクロホンアレイを用いて、音声、映像を同期させて撮影を行った。カメラ及びマイクロホンの配置図を、図2、図3に示す。カメラアレイは、半径4.5mの円周上に7.5度間隔に17台配置し、円の中心



図1 カメラアレイとマイクロホンアレイ  
Fig.1 Camera and microphone array.

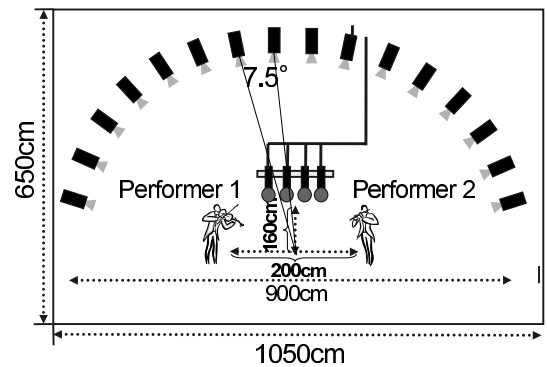


図2 カメラ配置  
Fig.2 Camera setup.

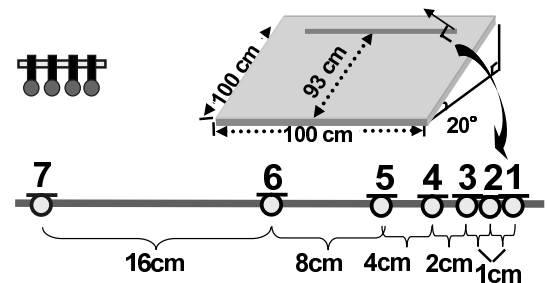


図3 マイクロホン配置  
Fig.3 Microphone setup.

を向く形で設置した．また，等比的な間隔で設置したマイクロホンアレーを円の中心 1 m 手前に設置した．

### 3. 撮影及び自由視聴点映像の生成

本システムを用い，楽器演奏を撮影した．2~3 人の演奏者は円の中心から半径 1 m 以内かつ，マイクロホンアレーから 1.6 m 手前に位置して演奏を行った．

収録された撮影データに対し，文献 [2], [3] の方法を適用することで，自由視聴点映像を生成した．映像・音声の両処理のフローを図 4 に示す．まずはじめに，カメラやマイクロホンの位置を測定するキャリブレーションを行う．次に，各画像の色合いやひずみ，マイクロホンの音量などを補正する．そして，シーンの奥行情報の計算や音源位置の特定を行ったあとで，自由視聴点映像を生成する．

文献 [2] は，Image Based Rendering に属する手法に基づいており，物体の奥行推定結果を用いて画像を三次元的にワーピングさせることで自由視点画像を生成する．自由視点画像の生成時間は約 200 ms である．文献 [3] は，周波数領域独立成分分析により音源分離をオフラインで行い，音源とユーザ間に対応する立体音響効果を付加することで，実時間で任意点の音像を定位させることが可能となる．今回は，両者の計算時間の差異のため，映像側の視点位置を限定し，その視点の画像データをオフラインで計算して統合した．入力となる 17 台のカメラ映像が滑らかにつながるように，もともと 7.5 度間隔で並んでいたカメラを 0.75 度間隔で撮影した映像となるように事前に計算した．映像位置は視点切替及び画像読み込み，音声はその視点位置での実時間自由聴点生成という条件下でインタラクティブなりアルタイム再生が可能となる．

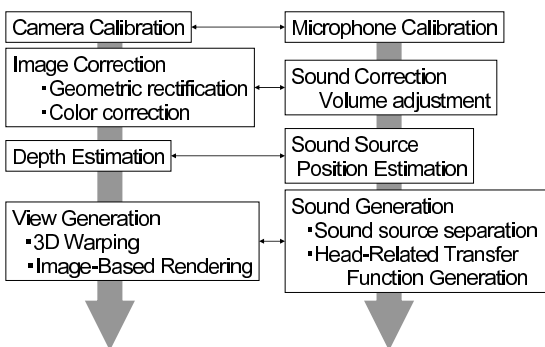


図 4 自由視聴点映像の処理フロー

Fig. 4 Flowchart of free viewpoint and listening-point video.

両者ともに確立されつつある技術ではあるが，二つのメディアを統合するために，各メディアで用いた時間軸と三次元座標を一致させる必要がある．前者は，音声映像の取得時に同期が完全にとれるシステムを用いたため解決される．また後者は，映像の三次元座標に音声の座標を合わせることで解決させる．実際には，音声・映像ともに別の基準をもって測量をするため，二つの座標系には差異が存在している．しかし，人間の空間認知能力を考慮すると，音声から三次元位置を特定する能力が，映像から特定する能力に対して著しく低いことから，両者の差異はほぼ影響がないと考えられるからである．立体音響効果に頭部伝達関数を用いた場合，音源定位に関する方向弁別限として，水平面前方で約  $1 \sim 2^\circ$  であることが報告されている [4]．一方，今回使用したカメラは水平方向の分解能は，約 0.02 度 (画角 29.5 度，解像度 1392 pixel) であり，再現する空間分解能のオーダが 2 けた異なる．そのため，音声側の座標をメジャーなどを用いて測定した物理的な値を用いてカメラで生成した三次元空間に合わせることで解決した．

### 4. 結果

自由視聴点映像の生成結果を示す．図 5 の左図は生成した自由視点画像である．また，その音声と同期したビデオは，ディスプレイとヘッドホンで再生可能であり，下記 URL [5] からダウンロード可能となっている．また，図 5 の右図は文献 [6] の円周立体ディスプレイへ生成した自由視点データを出力した結果である．今後は，これらと立体音源の出力を組み合わせることで臨場感のある再生を目指していく予定である．

### 5. むすび

本論文では，カメラアレー，マイクロホンアレーを用いて，三次元シーンの映像・音声を完全に同期して取得するシステムを構築した．それらを用いて，ユー

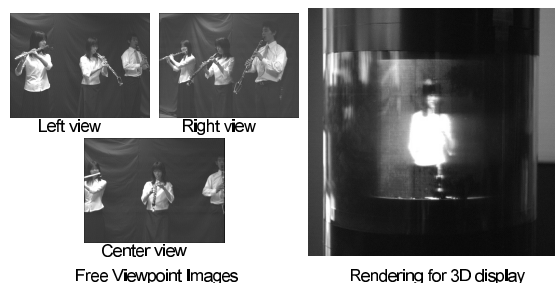


図 5 生成画像結果

Fig. 5 Rendering result.

ザが自由に視点や聴点を自由に選択可能となる自由視聴点映像生成を行った。自由聴点はリアルタイム再生可能であるが、自由視点がリアルタイム再生不可能であったため、再生する視点位置を限定しオフラインで事前に画像生成を行った。再生時には、映像は視点切換という条件下で、インタラクティブな再生が可能となる。

今後の予定として、今回の自由視聴点映像生成では音声・映像の各処理は独立に行ったが、同じ三次元シーンから取得される音声・映像間には大きく情報に相関性があるはずである。この相関性を使う、両データにわたる横断的な処理により、物体の位置推定やシーンの状況解析などの精度向上を目指す。また、このメディアの品質、臨場感などの評価手法を空間認識特性等を考慮しながら検討していく予定である。

## 文 献

- [1] B. Wilburn, N. Joshi, V. Vaish, E. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," Proc. ACM SIGGRAPH, July 2005.
- [2] N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto, "Free viewpoint image generation using MultiPass dynamic programming," Proc. SPIE SD&A XIV, vol.6490, pp.460-470, 2007.
- [3] 丹羽健太, 西野隆典, 武田一哉, "ブラインド音源分離と頭部伝達関数を用いた音楽演奏を題材とした自由視聴点コンテンツの制作," 信学技報, EA2007-32, 2007.
- [4] 黒澤 明, 都木 徹, 山口善司, "頭部伝達関数と方向弁別能力について," 音響誌, vol.38, no.3, pp.145-151, 1982.
- [5] <http://www.sp.m.is.nagoya-u.ac.jp/~niwa/fvlp.html>
- [6] 圓道知博, "周囲 360 度から観察できる 3 次元ディスプレイ Seelinder," 映情学誌, vol.59, pp.1506-1509, 2005. (平成 19 年 12 月 3 日受付, 20 年 3 月 14 日再受付)