

遠隔地のピアノ演奏ライブの実時間伝送とCGによる可視化

2M-10

- 遠隔地ピアノレッスンシステムの実現に向けて -

中山 綾子 興梠 正克 村岡 洋一

早稲田大学大学院 理工学研究科

1. はじめに

近年、映像と音声、音楽を効率良く伝送するための符号化手法として、モデルベース符号化が着目されている。映像対象に関する共通のモデル (eg. 3次元形状) を送信側と受信側であらかじめ持つておく。そして、そのモデルを制御するパラメータだけを伝送することで、非常に低いビットレートで高品質映像の送受信を可能にする。本研究では、人による鍵盤楽器 (ピアノ) 演奏の映像と音楽を対象として、演奏のモデルベース符号化・伝送と、受信側でその演奏の任意視点からの可視化を実現するシステムの構築を目的とする。

鍵盤楽器の演奏を可視化するためには、人間の演奏者の各指の運指とその軌道を正確に得る必要がある。関口らの研究 [3] では、ピアノ曲の楽譜情報のみに基づいて運指とその軌道を決定して、3DCG による演奏の可視化を試みていた。しかしながら、楽譜情報だけでは運指を一意に決められない状況も多く、また個人差による演奏方法の相違などを表現できないという制限があった。一方で、演奏者を撮影した映像からその手指の動作、軌道を決定する方法も考えられるが、映像中の物体の三次元形状や動きを推定する問題は一般的には不良設定であり、その解決は困難である。

そこで、本研究では、演奏者の手の動きを推定するための情報源として、演奏を撮影した映像とその演奏結果である楽音情報の両方に着目する。動画解析により各指の空間中の位置、移動方向と速度を推定する。その結果を、楽音情報に基づいて得られる情報 (各指の打鍵時刻、鍵盤位置とその強さ) と組み合わせることで、非剛体の動き推定問題を容易かつロバストにする。

本研究では、具体的に、MIDI 鍵盤楽器を演奏する人物を撮影した動画と、楽器から生成される MIDI 信号を入力とする。本システムでは、最初に、動画を解析して各指の空間中の位置と軌道をダイナミックに推定する。次に、MIDI 信号から打鍵された鍵盤位

置とその打鍵の強さの情報を得て、これを動画の解析結果と組み合わせる。その結果、より正確な両手の動作・姿勢パラメータを得ることができる。そして、これらのパラメータと MIDI 信号を符号化して伝送する。受信側では、両手の動作・姿勢パラメータに基づいて両手の多関節モデルの動きを復元して CG を用いて可視化する。CG モデルを用いて可視化することで、ユーザの望む任意視点からの演奏状況を再現する。

2. 本システムの概要

本システムは、以下の4つの処理過程からなる。

1. 動画解析 (送信側): 動画を解析して、各指の空間中の位置をダイナミックに推定する。
2. MIDI 信号解析 (送信側): MIDI 情報と動画解析結果を統合し運指パラメータを抽出する。
3. 符号化と伝送 (送信側): 符号化された運指パラメータと MIDI 信号を送信する。
4. 復号化と可視化 (受信側): 受信したパラメータを受け取り、CG により可視化する。

以下、1-4 の各処理について詳しく述べる。

2.1 動画解析

手指の位置を推定する動画解析を容易にするため、演奏者の各指に赤、緑、青色のマーカを装着しておき、映像の色 (HSV 色空間での色相値) を解析して、彩度が十分に高く、各画素の色相値がマーカの色と一致している画素領域を分離、併合してマーカ領域を背景から分離する。この領域の重心位置を求めることで、各フレームにおける各指の位置を推定する。

2.2 MIDI 信号解析

前節に述べた動画解析だけでは、各指の打鍵の有無とその強さ、継続時間を正確に推定することは困難である。そこで、映像から得られた各指の鍵盤上の位置と、MIDI 信号から得られた時刻 (タイムスタンプ)、打鍵された鍵盤 (note) とその強さ (velocity) を対応づけることで、映像中の各指が打鍵した鍵盤位置と打鍵の強さ、その正確な時刻を得ることができる。

Real-time model-based coding and CG visualization of human piano performance.

Ayako Nakayama, Masakatsu Kourogi, Yoichi Muraoka
School of Science and Engineering, WASEDA University

2.3 符号化と伝送レート

送信側では、演奏者の映像を復元できるように各時刻の各指の打鍵情報を表す運指パラメータを送信する。また楽音を復元できるように MIDI 鍵盤楽器の出力として得られる MIDI 信号を送信する。

本手法による伝送レートの条件を以下のように見積もる。まず、MIDI データの伝送量は 32Kbps (最大) である。次に、ピアノ演奏の速度として、四分音符=240 (1 分間に四分音符が 240 個) を上限と見なせる。鍵盤 256 個 (8bit), 指番号 1~10 (4bit), 32 分音符を 1 秒間に 32 個, 10 本の指で同時に打鍵したデータ量を最大にする状況において、 $(8 + 4) \times 32 \times 10 = 3,840$ bps \approx 4Kbps となる。したがって、伝送レートの上限は大きく見積もっても 36Kbps である。

2.4 復号化と可視化

受信側では、あらかじめ手指の (関節を含む) 3DCG モデルを作成、保持しておく。そして、MIDI 信号の再生と同期を取りながら、伝送された運指パラメータに基づいて 3DCG モデルを動作させる。また、この生成された CG モデルをユーザの指示に基づいて任意のズーム率と視点から可視化する。

3. 実装と評価実験

本システムを以下の計算機環境に実装した。動画解析と MIDI 信号解析に基づく運指パラメータ抽出には SGI Octane ワークステーション (R10000 195MHz) を用いて、Video Library (VL) で画像を取り込み、MIDI Library (ML) で MIDI 信号の取り込みを行った。パラメータと MIDI データの伝送にはソケットによる TCP/IP 通信により実装した。CG による演奏の可視化は、SGI Onyx ワークステーション (R10000 250MHz) を用いて、OpenGL ライブラリにより実装した。

前述の計算機環境上に実装された本システムにより、MIDI 鍵盤楽器を演奏する人物を撮影した映像と MIDI 信号を入力として、その処理速度と伝送レートについて評価する実験をおこなった。その結果、(1) 運指パラメータの抽出、(2) これらのパラメータと MIDI 信号の伝送、(3) CG による可視化の実時間処理に成功した。各フレーム画像の処理に要した計算コストは、動画解析に 20 ミリ秒、MIDI 信号との同期と対応づけに 5 ミリ秒であり、符号化には 1 ミリ秒以下であり、処理の実時間性を確認できた。また、1 曲の平均的な伝送レートは 8Kbps 程度であることが確認できた。本システムの処理の全体像を図 1 に示した。

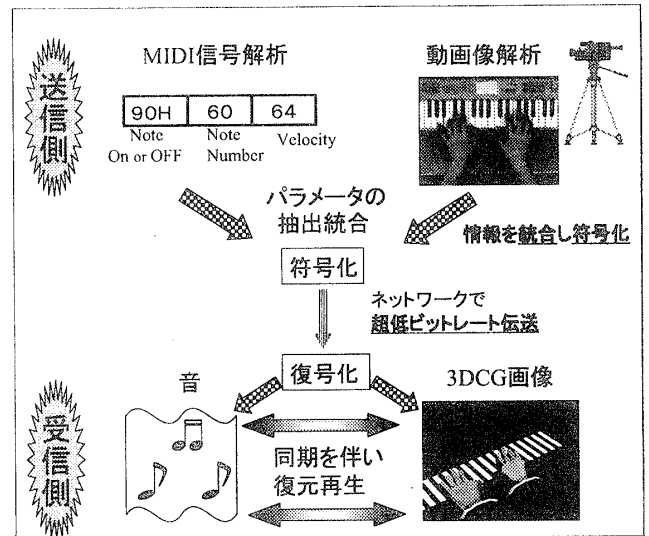


図 1: 本システムの全体図

4. おわりに

鍵盤楽器の演奏を超低ビットレート (10~30Kbps) で実時間伝送することをモデルベース符号化を用いることにより実現した。また、画像解析・MIDI 信号解析を統合することにより安定したパラメータを抽出できた。さらに、CG を用いることにより、任意の視点からの実時間の再現を実現できた。

今後はよりリアルな再現を行うために、指の多関節モデルと各指の打鍵状態を合わせて各関節角度の逆推定を実現する。さらに、滑らかな表現を行うために、ある点から次の点への形状間に関節角度の補間を行う。

参考文献

- [1] "MPEG-4 visual committee draft 14496-3", ISO/IEC/JTC1/SC29/WG11 Document 1997.
- [2] H. Murase, S. K. Nayar, "Visual learning and recognition of 3-D objects from appearance", IEEE Trans. on Image Processing, Vol. 28, No. 3, pp.421-430, 1995.
- [3] 関口, 英保: "計算機によるピアノ演奏動作の生成と表示," 情報論文誌, Vol. 40, No. 6, June, pp.2827-2837, 1999.