研究報告

質点の物理モデルを用いた動画像からのソニフィケーション
SONIFICATION FROM THE VIDEO USING PHYSICAL MODEL OF A MATERIAL POINT

原 一歩
Kazuho HARA
東京電機大学
Tokyo Denki University

小坂 直敏
Naotoshi OSAKA
東京電機大学
Tokyo Denki University

概要

筆者らは，背景と異なる色を持つ物体の動きのある動画像を対象とし，これと同期する音響を付与する，動画像からのソニフィケーションシステムを構築している．このシステムは，まず，同画像内の物体の移動を質点運動ととらえ，フレーム間差分により速度と加速度を算出する．ここで，フレーム画像から追跡対象の物体以外の情報を除去し，重心点を算出し，これを質点の座標とした．次に，加速度から質点の衝突判定を行い，衝突判定時にリアルタイムに FM 合成する．この方法を球と人の動きに適用して，その効果の有効性を確認した．

1. はじめに

一般に，映像作品は映像に同期して音響を付与することにより，視聴者により効果的な印象を与える．しかし，映像内のある物体の動作に合わせて音響を付与する作業の負担は大きい．森下らは動画像内の物体を追跡し，入力した衝突音を，物体の移動に合わせて半自動的に付与するシステムを構築した [1]．しかし，このシステムは，物体の移動に対応した，衝突音の自動的な音量の変化が行われており，効果音の音色の変更は行わない．また，システムは，入力された動画像にリアルタイムに音響を付与しない．そこで，本稿では，ソニフィケーションの対象の動画像を，背景と異なる色を持つ物体が 1 つだけ移動しており，30fps であるものに限定し，以下の 2 点を満たすシステムを構築する．1) 入力された動画内の物体の動きを質点運動と捉え，加速度から衝突判定を行う．2) 衝突判定時に，加速度の大きさを音合成のパラメータとして入力し，リアルタイムに FM 合成を行う．本稿は構築したシステムの評価実験を行い，物体追跡の精度を測定する．また，システムを様々な動画に応用し，その実用性を考察する．

2. 物体追跡と衝突判定

本稿では，フレーム間差分と HSV 色空間による色抽出を行い，動画像内を移動する特定の色の物体を各フレーム画像から抽出する．次に，抽出後のフレーム画像の重心を計算し，質点の座標とする．

2.1. 色重心を用いた動画像の物体追跡

まず，動画像の各フレーム画像と，そのフレーム画像を基準とした時の，過去の 2 フレームの画像をグレースケールに変換し，フレーム間の差分画像を求め，2 値化する．次に，2 回のオープニング処理を行うことにより雑音を除去する．これにより，各フレーム画像から移動している物体を抽出する．

移動している物体が抽出された 2 値化画像と同フレーム番号における RGB 画像で論理積を取り，移動している物体の RGB 画像を抽出し，各画素値を HSV 色空間に置き換える．最後に HSV 色空間において閾値の判定を行い，色に収まっているかどうかで 2 値化することにより，フレーム画像から移動する特定の色の物体を抽出する．画像処理後の各フレームの 2 値化画像において，縦方向と横方向でそれぞれヒストグラムを算出し，平均を求めることにより重心の座標を求める．次に，各座標値を 0〜100 に正規化し，質点の画面座標とする．

2.2. 質点の加速度からの衝突判定

各フレーム画像における質点の座標と，そのフレームを基準として，過去の 2 枚のフレーム画像の質点の

座標を，前後のフレーム同士で差を取り，質点の速度を 2 つ算出する．さらに，速度ベクトルの差を取り，各フレーム画像における加速度とする．衝突時の力に比例する量として，二次元加速度ベクトルの絶対値を用いて，絶対値が指定された閾値を越えた時のフレーム番号において，物体の衝突が発生したと判定する．

3. 音合成

衝突判定時に，加速度を任意に定数倍した値を，周波数変調 (FM: Frequency Modulation) のキャリア周波数と変調周波数および変調指数に代入し，式 (1) により音波形を合成する．また，FM による合成音波形を減衰させるためにエンベロープをかけ，これを衝突音とする．

$$y(t) = Asin(2\pi f_c t + Isin(2\pi f_m t)) \qquad (1)$$

ここで、t: 時刻 [sec]，f_c: キャリア周波数 [Hz]，f_m: 変調周波数 [Hz]，I: 変調指数，A: 振幅，$y_n(t)$: 合成波形である．

4. システムの構成

ソニフィケーションシステムの実装に使用したツールを表 1 に示す．Pure Data[2] は Miller Puckette により開発された，マルチメディアの表現に特化したビジュアルプログラミング言語である．OpenCV[3] は Intel により開発された，画像処理のライブラリ群である．OpenSound Control[4](以下，OSC と略す) は，M. Write が開発した，音響関係のデータの送受信に特化した通信プロトコルである．本稿では，物体追跡の処理と，動画像再生に同期した音合成の処理にシステムを分割し，同一のマシン内で実装を行った．各処理部のシステム構成については，4.1 章および 4.2 章で説明する．

表 **1**. 使用ツールの一覧

ツール名	用途
C++	質点の画面座標からの加速度の算出
OpenCV	画像処理による物体追跡
Pure Data	加速度に対応した音響信号の合成
OSC	プログラム間の加速度情報の送受信

4.1. 物体追跡部のシステム構成

物体追跡部のシステム構成を図 1 に示す．システムには動画像と，HSV 色空間における抽出したい色の範囲設定のデータを入力するようにした．各画像処理の実装は，OpenCV のライブラリを用いて行った．画像処理後の各フレームのそれぞれの重心点から，フレーム毎の加速度を計算し，リストファイルとして外部出力する．

4.2. 動画再生及び音合成のシステム構成

動画再生と音響信号の合成のシステム構成を図 2 に示す．動画再生部は C++により実装し，音響合成部は Pure Data を用いて実装した．また, OSC を用いて, C++ プログラムによる動画再生と，Pure Data プログラムによる音響信号の合成を同期させた．図 2 のシステムには，動画像と，それに対応するフレーム毎の加速度のリストファイルを入力として与える．

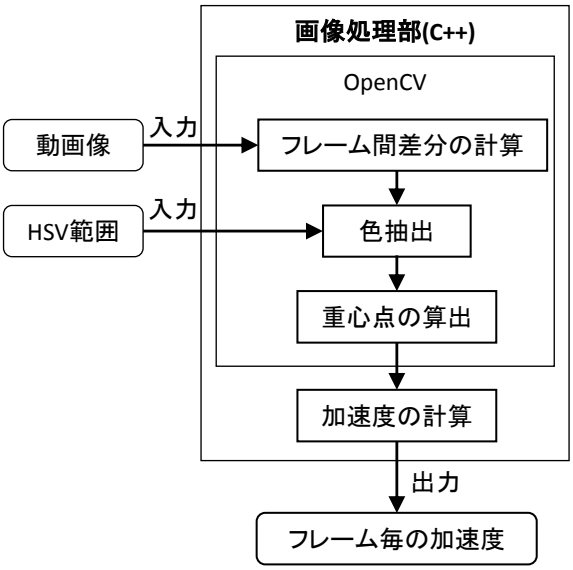

図 **1**. 物体追跡部のシステム構成

5. 評価実験

本システムの物体追跡と衝突判定の精度を，背景やカメラの環境毎に調査するため，評価実験を行った．

5.1. 実験手法

まず，表 2 に示す要因と水準に基づいた，それぞれの環境下で色のある球体を落下させ，床に衝突して跳ね返る動画を撮影し，実験データとする．実験に使用した背景画像 [5] を図 3 に示す．次に，実験データの動画像から目視できる衝突のフレーム番号の一覧を作成する．また，システムに動画像を入力して得られる加速度が，閾値を超えた時のフレーム番号の一覧を用意する．2 つの一覧を照合し，前後 1 フレームの誤差

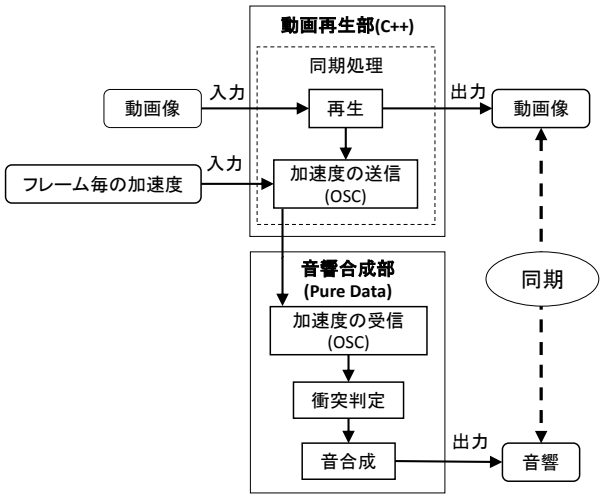

図 **2**. 動画再生及び音合成のシステム構成

内であれば，システムが正しく判定した衝突フレームとする．最後に，適合率と再現率および F 尺度の各評価を，式 (2)，(3) から算出し，実験結果とする．今回の実験では，動的特徴量の閾値を 1 に設定した．

$$p = \frac{R}{N}, r = \frac{R}{C} \tag{2}$$

$$F = \frac{2pr}{p+r} \tag{3}$$

ここで、p: 適合率，r: 再現率，F: F 尺度，R: システムが正しく判定した衝突のフレームの枚数，N: システムが判定した衝突フレームの枚数，C: 目視できる衝突フレームの枚数である．

表 **2**. 実験データの要因と水準

要因	水準			
背景画像	有	無		
カメラの高さ	0m	1m		
カメラの俯瞰角度	0°	45°		
ボールの色	赤	青	黄	白

5.2. 実験結果と考察

実験により得られた動画像毎の評価値を，背景画像の有無毎に集計したものを表 3 に示す．背景画像がない場合の F 尺度は 0.667 であり，この性能は厳密に質点の動作と対応させる目的では満足できないが，全体のグローバルな動きに音を与える意味では許容できると考えた．特に，人の体の重心のように手足の個別の動きではなく，それら複合体の動き全体に同期させる

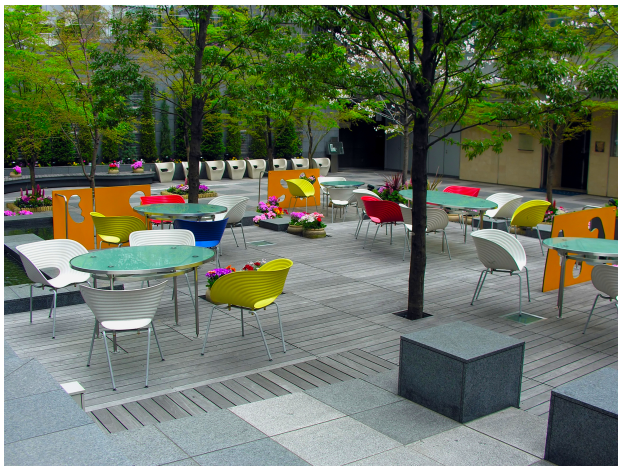

図 **3**. 実験で使用した背景画像

場合は，音と対応させる対象がまぎれてしまうため，この性能でも十分効果がある，と考えられる．また，背景画像がある場合の F 尺度は 0.690 であり，背景画像の有無において F 尺度の差異は見られなかった．このことから，背景に動きがない場合，背景画像がない場合と同程度の性能で物体追跡が行われていることがわかった．

次に，動画像毎の評価値をカメラの俯瞰角度毎に集計したものを表 4 に示す．同表の F 尺度を見ると，俯瞰角度を 45° に設定した場合の F 尺度は 0.723 であり，俯瞰角度を 0° にして撮影した場合より高い値を示している．これは，落下する球体を水平に撮影すると，動画内における物体の移動速度が早くなり，残像による球体の色調の変化により，正常な物体追跡が行えなかったからであると考えられる．

表 **3**. 背景画像の有無毎の評価値

背景画像	適合率	再現率	F 尺度
有	0.719	0.663	0.690
無	0.687	0.649	0.667

表 **4**. カメラの俯瞰角度毎の評価値

カメラの俯瞰角度	適合率	再現率	F 尺度
0°	0.706	0.563	0.627
45°	0.699	0.749	0.723

6. システムの応用と考察

構築した本システムを用いて，球体あるいは人間の運動する動画像からのソニフィケーションを試み，現

段階のシステムの実用性を考察する．表 5 に入力として与えた動画像の内容を示す．また，同表の各データから抜粋したフレーム画像と，それに対して追跡する物体の抽出を行った結果をそれぞれ図 4，5 に示す．データ#2 では，図 5a のフレーム画像内の，赤色のパーカーを追従することにより，跳躍した人間が床に着地した時に対応した音響の合成を試みる．

表 **5**. 入力する動画像データの内容

データ	内容	追跡対象
#1	球体の跳ね返し合い	桃色の球体
#2	人間の跳躍（スキップ）	赤色のパーカー

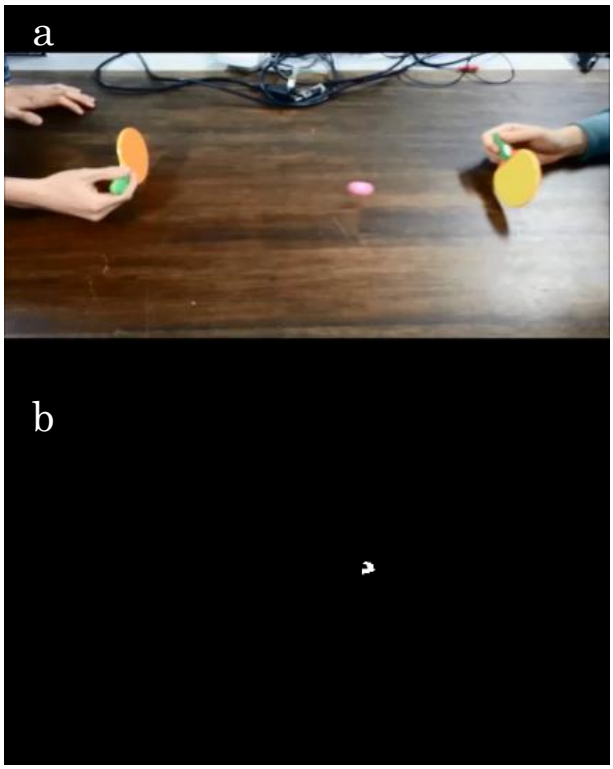

図 **4**. データ#1 のキャプチャ画像 a と物体抽出の結果画像 b

6.1. データ**#1** のソニフィケーション結果と考察

図 4a のフレーム画像では，追跡する球体と似た色味を持つ人間の手が映っている．しかし，図 4b を見ると，HSV における色抽出により，フレーム画像から追跡対象の球体のみ抽出できている．また，球体がラケットにより弾かれた時に，概ね衝突判定がされ，タイミングに合わせて音響の付与が行われた．これは，5.2 の実験に比べ，衝突がより抽出しやすい条件で行ったためと考えられる．音響に関しては，ラケットに衝突した球体の速度の変化に対応し，動画の演出を向上するような音響が合成されていると考えられる．

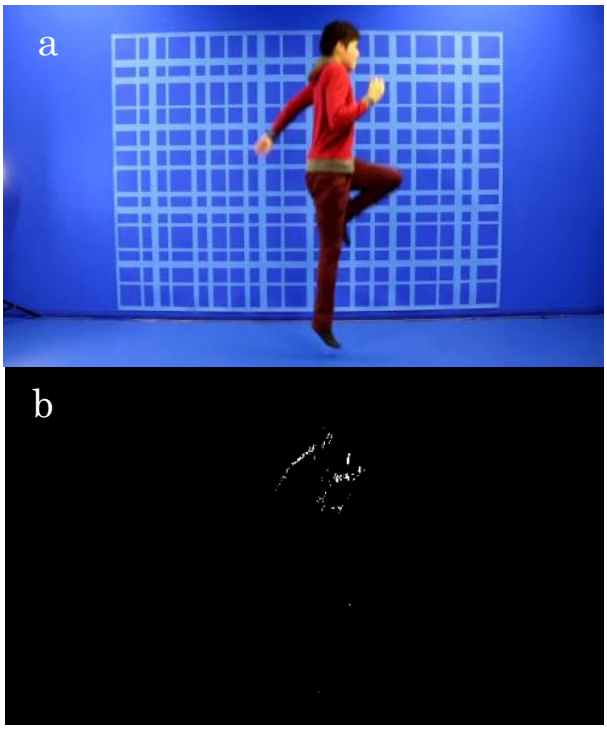

図 **5**. データ#2 のキャプチャ画像 a と物体抽出の結果画像 b

6.2. データ**#2** のソニフィケーション結果と考察

図 5a のフレーム画像から追跡する物体の抽出を行った結果，赤色のパーカーだけ抽出することができた．しかし，図 5b の抽出結果を見ると，元のフレーム画像の赤色のパーカーが映っている箇所のうち，一部しか抽出が行われていなかった．そのため，重心点の計算により求められる質点の座標の精度が低下し，動画内の人間の着地以外の場面において，衝突の判定が多発した．これは，赤色のパーカーの画面内の移動速度が，その色味を有している画素の面積の広さに対して遅いため，フレーム間差分の算出時に，除去されてしまったからであると考えられる．

7. まとめ

本稿では，動画像内の物体の動作から質点の加速度を算出し，衝突判定時に音響をリアルタイム合成するシステムを構築した．衝突判定の物理評価実験の結果，背景画像が動いていない場合は，背景画像がない場合

と同程度の性能で物体の追跡が行えることがわかった．一方，移動速度の高い球体を撮影すると，各フレーム画像において球体の被写体ぶれが生じ，正常な物体追跡が行えなかったことも明らかになった．また，動画像内の追跡する対象物を球体以外にし，動画像からのソニフィケーションを行った結果，質点の座標の精度が低下し，衝突の誤判定が多発した．以上の課題点を改善するために，より高度な画像処理を導入することを検討している．

8. 参考文献

[1] 森下沙耶, 岡部誠, & 尾内理紀夫. (2012). 動画への効果音付加支援システムの作成 (学生研究発表会). 映像情報メディア学会技術報告, 36(8), 119-122.

[2] Puckette, M. (1996). Pure Data: another integrated computer music environment. Proceedings of the Second Intercollege Computer Music Concerts, 37-41.

[3] OpenCV http://opencv.jp/

[4] Wright, M., Freed, A., & Momeni, A. (2003, May). Opensound control: State of the art 2003. In Proceedings of the 2003 conference on NIME(New Interfaces for Musical Expression) (pp. 153-160). National University of Singapore.

[5] 藤田紘久 Futta.NET http://www.futta.net

9. 著者プロフィール

原 一歩 (Kazuho HARA)

東京電機大学大学院未来科学研究科情報メディア学専攻音メディア表現研究室所属．高校 2 年から DTM による作曲を行っており，大学院入学後にピアノを始める．現在は，動画と静止画からのソニフィケーションに関する研究を行っている．

小坂 直敏 (Naotoshi OSAKA)

昭 51 早大・理工・電気卒．昭 53 同大大学院修士課程了．同年日本電信電話公 社 (現 NTT) 入社．以来通話品質の研究，音声対話の研究，コンピュータ音楽あるいはマルチ メディア創作のための音響研究などに従事．平 6 早大より博士 (工学)．平 8-14 コミュニケーション科学基礎研究所音表現およびメディア表現 研究グループリーダ, 平成 15 東京電機大学 工学部教授．メディアコンテンツのための音響情報処理の教育と研究に従事．また，音 楽制作および発表活動も行う．2006 に開始した Media Project 他，コンピュータ音楽のコンサート企画多数．日本音響学会，電子情報 通信学会，情報処理学会，ICMA, IEEE 日本電子音楽協会各会員．現在，東京電機大学 未来科学部教授．本会会長．