

動画像の色重心とその動的特徴量を用いたソニフィケーション

原 一步^{†1} 上野 崇広^{†1} 小坂 直敏^{†1}

この論文は、入力された動画像の色調の変化とオブジェクトの動きに対応した音をリアルタイムに出力するソニフィケーションシステムについて述べたものである。システムはビジュアルプログラミング言語である Pure Data を用いて構築した。RGB のヒストグラムから求めた色の平均とフレーム間での重心の移動距離との積を動画における特徴量とした。特徴量をピッチ周波数に割り当て、動画の変化に対応した音響、及び音楽を出力した。

1. はじめに

時間と共に変動する科学データや音響情報以外の情報を、音響信号に変換する表現を可聴化、またはソニフィケーションという。この話題自体が一つの学会となり 92 年以降様々な研究がなされてきた⁽¹⁾。本稿では動画像からのソニフィケーションを対象とする。一般に、動画と音声は互いに相補的であり、強調、または対比など関係を持って、映像作品の演出をより向上させている。本研究のねらいは、ソニフィケーションを用いて、入力された動画から特徴量を抽出し、ピッチ周波数に割り当てることにより、動きに対応した音響を出力するシステムを構築することである。

2. 動画からの特徴量の抽出

動画情報をソニフィケーションするにあたり、合成音にも動画の持つ時間的な構造を保持させることとする。

具体的には以下の 2 点を満たす。1) 音色が同画像の時間的な変化に追従し、急激な変化には出力音も急激となるように対応させる。2) あるオブジェクトが動くとき、出力音も変化するようにする。

1) では、画面の中に花火が広がった時に音色が変化する例が考えられる。

一方、2) の例として、画面の中で人やボールなどのオブジェクトが動いているような動画がある。これは、ヒストグラムを特徴量とした変換ではあまり変化しないため、音色も変化しない。そこで、色重心の単位時間 (1 フレーム) あたりの移動距離を特徴量の 1 つとした。

以上の 2 つの特徴量を各 RGB 毎に乗ずることによって、新たな画面の色調の変化と物体の動きに対応した特徴量を定義した。以下詳細に述べる。

2.1 色の平均の算出

33ms 毎のフレームから各 RGB 毎に 0 から 255 の画素値のヒストグラムをとり、ヒストグラムから求まる平均 $\bar{C}$ を特徴量の 1 つとした。

2.2 色重心の算出

まず、動画の現在のフレームから色重心を求めるために、色の明るさを重みとして、式 (1) を用いて色重心の画素単位の位置を計算した。

$$\vec{G} = \frac{\sum_{i=1}^N m_i \vec{x}_i}{\sum_{i=1}^N m_i} \cdot \dots \cdot (1)$$

$\vec{G}$: 色重心の座標

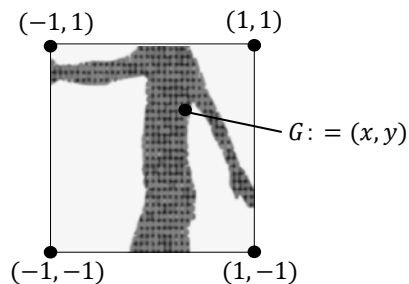
N : 全画素数

m_i : 1 つの画素の色の明るさ ($0 \leq m_i \leq 255$)

$\vec{x}_i$: 1 つの画素の座標

i : 画素の番号

本方式では、式 (1) を用いて求めた画素単位の色重心の座標を、図 1 のように、画面の右上端が (1,1)、画面の左下端が (-1,-1) となるように正規化した。



G : 動画のフレームの色重心

x : 色重心の横方向の座標 ($-1 \leq x \leq 1$)

y : 色重心の縦方向の座標 ($-1 \leq y \leq 1$)

図 1 画面座標上での重心点の算出

2.3 色重心の移動距離

30fps の動画に対して一つ前のフレームと現在のフレームで色重心の座標を求めた後、フレーム間で 2 点の色重心点間を結ぶ線分の長さを求めることによって、フレーム間の移動距離を求めた。これを RGB 全てに対して行う。

^{†1} 東京電機大学

2.4 最終的な特徴量の算出

2.1～2.3 からヒストグラムの平均とフレーム間の色重心の移動距離を取得することができる。次に、これらの積を各 RGB 毎にとることによって、色調の変化とオブジェクトの画面内での動きに対応した特徴量を (2) 式のように算出した。

$$p = \bar{C} \times \Delta X \cdots (2)$$

p : 色の平均と重心の移動距離の積

$\bar{C}$: ヒストグラムの平均

ΔX : 色の重心のフレーム間の移動距離

3. 音の出力

動画の画面から算出された特徴量と音響のピッチ周波数を対応付けするため、動画の画面から抽出された特徴量のうち、0 から 0.1 までの範囲内にあるものを、0Hz から 2kHz までのピッチ周波数へ線形にマッピングした。マッピングの例は以下の図 2 の通りである。実際は f_c はピッチ周波数であり、次節で述べる FM のキャリア周波数である。

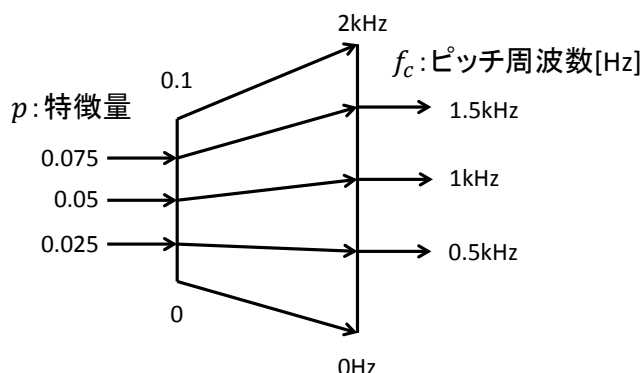


図 2 特徴量からピッチ周波数へのマッピングの例

3.1 周波数変調による音合成

ピッチ周波数を用いて周波数変調 (FM; Frequency Module) による音合成を行った。出力される音響に高次倍音が付加される。本システムでは変調周波数をキャリア周波数の整数倍に設定した。これにより、FM のキャリア信号の波形関数は、正弦波、矩形波、三角波、ノコギリ波の 4 種類から選択できるようにした。

各 RGB 毎に生成された波形は、最後に加算して 1 つの波形に合成し、サウンドの出力デバイスへ送信した。

4. システムの構成

4.1 使用プログラミング環境

システムは、ビジュアルプログラミング言語である Pure Data (Pd) を用いて構築した。Pd はビジュアルプログラミング言語であり、パッチやオブジェクトを組み合わせるリアルタイムに音声信号や動画を処理することができる。本研究では、プログラミングで用いられる頻度の高いライブラリが予め組み込まれた Pd-extended 版を使用した。

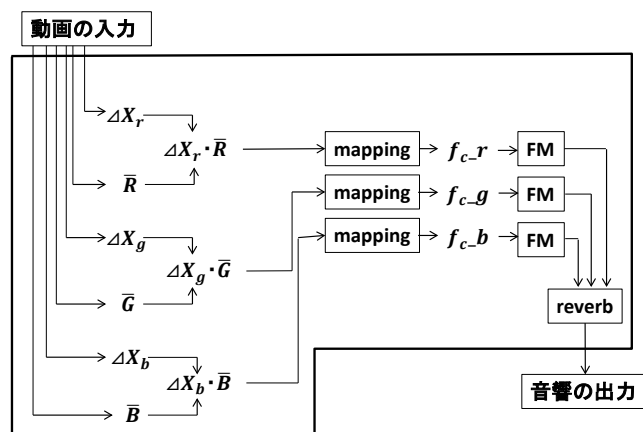
Pd は External と呼ばれる追加プログラムを用いることによって、C 言語による独自の拡張パッチの作成ができる。Pd だけでは実装できないような処理を、External を用いることによって実装した。External に使用したソフトを下表に示す。

表 1 External に使用したソフトの一覧

ソフト名	用途
Visual Studio C++ 2010 Express	C 言語によるソースコードの記述
CMake	Visual Studio のプロジェクトファイルの作成

4.2 システムの構成図

システムの全体の構成図を、図 3 に示す。



$\Delta X_r, \Delta X_g, \Delta X_b$: 色重心のフレーム間の移動距離

$\bar{R}, \bar{G}, \bar{B}$: 色重心のフレーム間の移動距離

$f_{c-r}, f_{c-g}, f_{c-b}$: 各RGBのピッチ周波数

mapping : 特徴量からピッチ周波数へのアサイン

FM : 周波数変調方式による音響信号の発生

reverb : リバーブの付与

図 3 ソニフィケーションにおけるシステム構成図

4.3 音階に対応したピッチ周波数の割り当て

アサインされたピッチ周波数を五音音階もしくは全音音階に対応したピッチ周波数に割り当てた。特徴量を音階

に対応したピッチ周波数へ変換される様子の一例を、図 4 に示す。この例では、特徴量から 349Hz に割り当てられ、五音音階もしくは全音音階に対応した周波数に再度割り当てられる。

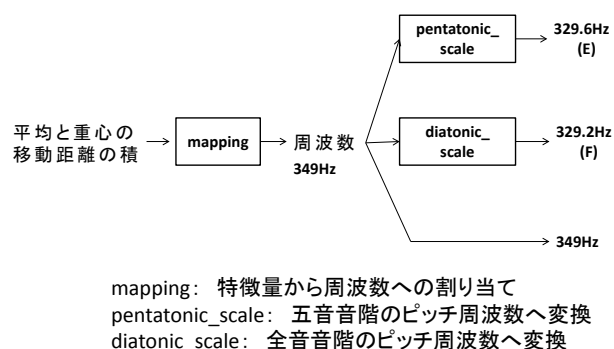


図 4 特徴量からピッチ周波数へ変換する一例

4.4 反響音の付与

出力される音に反響音を付加するため、波形に対し、ディレイのエフェクトをかけた。パラメータは入力された信号を出力するまでの時間と、出力される音響信号の振幅の 2 つとした。エフェクトからの波データを入力した波データに加算して合成することによって、出力する音にディレイがかかる。また、その波データをエフェクタにフィードバックすることで、反響音を付与することができる。

5. 入力動画からの出力結果

制作したシステムにサンプルの動画を読み込ませ、リアルタイムに音声出力させた。今回使用した 4 種類の動画を表 2 に示す。使用した動画の 1 フレーム分の画像を図 5 と図 6 に示す。また、それぞれの動画に対して結果を表 3 に示す。総じて、2 章で述べた 2 つの効果は得られている。

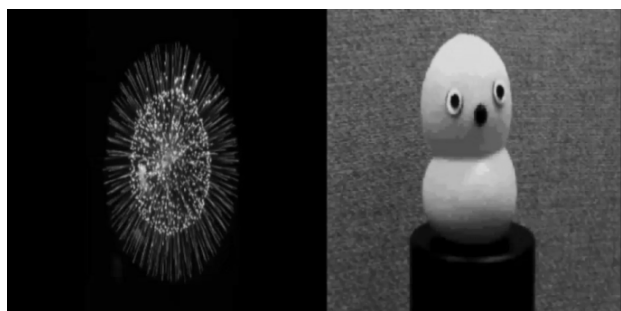


図 5 入力動画の 1 フレーム①

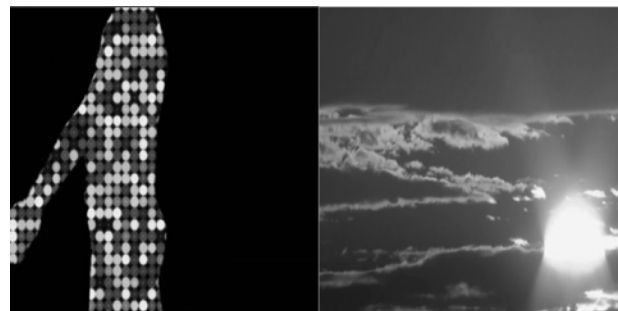


図 6 入力動画の 1 フレーム②

表 2 入力した動画の内容と特徴

データ	タイトル	内容	特徴
#1	花火	花火が開く	色調の変化
#2	ヒヨコ	ヒヨコが左右に揺れる	色重心の移動
#3	シルエット	右から左へシルエットが動く	色重心の移動
#4	夕焼け	赤みがかった夕焼け空	赤の明るさが高い

表 3 入力した動画と出力された音響の結果

データ	ソニフィケーションの適用結果
#1	花火の爆発と音色の変化が対応した。
#2	ヒヨコの左右の揺れに対応して音色が変化した。
#3	シルエットが現れている間だけ音響が出力された。
#4	音色に変化が見られなかった。

6. まとめ

ソニフィケーションを動画像に対して行い、動画像の変化に対応して音声を変化させながら出力するシステムの構築を行った。本研究では動画像の色重心の移動距離と、色のヒストグラムから求めた平均値から特徴量を計算し、特徴量からピッチ周波数へとアサインした。FM 音源を用いて、音響を出力した。

上述した過程を経て、動画像に対応した音声をリアルタイムに出力することができた。

問題点は、色の明暗に応じて、出力する音や音楽の明暗の変化が少ないことである。特に、花火の動画では、花火が爆発した時の音の変化に鮮明性が少なかったように感じた。動画像の動きから出力される音の変化をよりメリハリを付けるようなシステムを実装することが課題である。

本研究では、出力される音に対して、FM による変調や

反響エフェクト，オシレータによる音色の変化を与えることにより表現の幅を広くした．それに加えて，新しい特徴量を動画から抽出し，そこから得られる新しい音響効果を逐次システムに加える事によって，音の表現がより広がるシステムの構築が期待される．

謝辞 本研究を進めるにあたり，日頃熱心な議論をして頂く研究室の皆様に深く感謝します．

参考文献

- 1) <http://www.icad.org/biblio>
- 2) 中野稜允，大内誠，清野明日美，赤井澤博貴，小室和史，岩谷幸雄，関善一，久米祐一郎，“バーチャル聴覚ディスプレイを応用した視覚障害者の移動音源定位能力訓練の研究”，2013, 2
- 3) Miller Puckette: “PureData”, <http://puredata.info/>
- 4) cycling74: “Max”, <http://cycling74.com/products/max/>
- 5) Microsoft: “Visual Studio C++ 2010 Express”, <https://www.microsoft.com/ja-jp/dev/2010/overview/default.aspx>
- 6) Andy Cedilnik, Bill Hoffman, Brad King, Ken Martin, Alexander Neundorff: “CMake”, <http://www.cmake.org/>
- 7) Pure Data 入門 - こじ研（音・映像メディア）
<http://www.myu.ac.jp/~xkozima/lab/avmed-pd1.html>
- 8) ES... (Molojun): 動画素材.com, <http://xn--hhro09bn9j8uh.com/>