無音動画に対する効果音貼り付けシステムの検討

鈴木喜也　　岡部誠　　尾内理紀夫*

概要. 本論文はこれまで我々が研究してきた，無音動画に対し自動で効果音合成を行うシステムを改良するものである．このシステムは教師データとなる音付き動画を与え，その音データを切り貼りすることにより無音動画に適した新しい音データを生成する．先行研究の時点ではアルゴリズムの性質上，音の繋ぎ目が目立つ，連続性が保たれないといった問題があった．そこで本論文では，音付き動画から効果音の合成における特徴量と音データを必要な部分だけ切り出し，オブジェクト化するようシステムを改良した．この結果，より自然な音の合成が可能になる他，効果音の差し替えや加工も可能となった．

1　はじめに

近年，ニコニコ動画や YouTube といった動画コンテンツの普及が進んでいる．それと同時に動画製作ツールもフリーウェア，シェアウェア問わず増加しているが，効果音の合成に重点をおいたソフトウェアは少なく，効果音の合成にかかる労力は未だ大きいままというのが現状である．そこで我々は効果音合成の工程を効率化することによる動画製作における負担の軽減を考えている．

本論文では，先行研究において開発した効果音の自動貼り付けシステムにおけるアルゴリズムの改良を行う．改良後のアルゴリズムは物体の動きの一部を切り抜くことにより，動きのデータをオブジェクト化して扱うため音の差し替えや動的な音の編集を可能にする．

2　先行研究

動画に対する効果音合成の研究は物理演算を用いたものが多い [3]，[2]．また物理演算を用いない効果音の自動合成の研究として Cardle らによる研究 [1] が挙げられるが，この研究における手法は動画内の物体のモーションキャプチャデータが必要であり，これを一般の動画制作の場において使用することは難しい．

我々は昨年度の研究において物理演算を使用せず，モーションキャプチャデータも不要な効果音の合成システムの試作を行った [4]．試作したシステムは動画の各フレーム画像を解析し，画像から動画内の物体の動きを特徴量として抽出，既存の動画の音を用いて効果音の合成を行う．しかしこのシステム内で使用しているアルゴリズムは効果音を細切れにして新しい音データを生成するため，音の途切れやノイズが発生しやすいという問題があった．

そこで本論文ではこのシステムを改良し，教師データから必要な部分のみを抽出するようなユーザーインタフェースを作成し，効果音の質の改善を行った．

3　システム概要

図 1 にシステムの処理フローを示す．システムは以下の順序で効果音の合成処理を行う．赤字で示した部分が本論文における改良において追加された処理である．

1. ユーザーが事前に作成した無音動画と既存の音付き動画を入力として与える．
2. システムは入力された双方の動画の各フレーム画像から特徴量を抽出し，外部ファイルに保存する．
3. 2 において音付き動画から抽出された特徴量と音の対応から，効果音の合成に必要となる部分の抽出を行う．抽出された特徴量とそれに対応した音データは動作オブジェクトとして外部ファイルに保存する．
4. ユーザーは 3 において抽出された動作オブジェクト内の音データの編集を行う．
5. 3 において抽出された動作オブジェクトがもつ特徴量と，1 において無音動画から抽出された特徴量のマッチングを行う．マッチングの結果として，音データを挿入するフレーム番号のリストを作成する．
6. 5 において得られたフレーム番号のリストを元に，動作オブジェクトがもつ音データを適当な位置に挿入し，新しい音データを得る．

* Nobuya Suzuki, 電気通信大学, Makoto Okabe, 電気通信大学/JST PRESTO, Rikio Onai, 電気通信大学

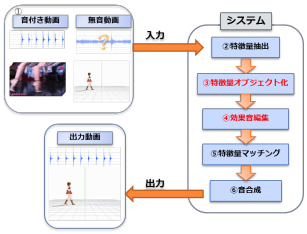

図 1. システムの処理フロー

図 2. 特徴量と音データの提示

4 改良点

4.1 特徴量の切り出し

システムは特徴量として入力動画から抽出したオプティカルフローを用いる．これまでのシステムは，音付き動画から抽出した全ての特徴量を使用していたが，実際に効果音の合成時において使用する特徴量はその中の一部であり，それ以外の部分は合成におけるノイズになっている場合がほとんどであった．

本論文では合成に使用する特徴量と音データを持った，動画とは独立したオブジェクトを作成する．このオブジェクトは動作オブジェクトと呼ぶ．

システムは特徴量の解析が終わると図 2 のようなウィンドウを表示し，ユーザーに対して抽出した特徴量と音データの提示を行う．ユーザーはそれらを見て必要な部分をドラッグし，選択する．システムはドラッグされた範囲内に存在する音データともう一方のデータの対応する部分を抽出し，新しく作成した動作オブジェクトに格納する．

4.2 特徴量マッチング

第 4.1 節において得られた特徴量データを用いたマッチング処理について説明する．

1. 無音動画と動作オブジェクトの各オプティカルフローの値を，各データのオプティカルフローの L2 ノルムの最大値が 1 になるよう正規化する．
2. 無音動画の特徴量データを時間軸に沿って動作オブジェクトの特徴量と同じサイズ分切り出す．
3. 2 で取得した無音動画の特徴量と動作オブジェクトの特徴量に関して，オプティカルフローの値のユークリッド距離の総和を算出する．
4. 3 で算出したユークリッド距離の総和が ε 以下ならば 5 に進む．そうでなければ 2 に戻り処理を繰り返す． ε は音の合成を行うかどうかの閾値であり，式 (1) により算出する．

$$\epsilon = \rho \sum_{i=1}^{N} L(x_i, y_i) \qquad (1)$$

L は各動作オブジェクト内のオプティカルフローの L2 ノルム， ρ はユーザーが任意に設定する倍率であり， $\rho = [0,\ 1]$ である．現在は暫定的に $\rho = 0.3$ として実験を行っている．

5. 4 の条件を満たした位置に，動作オブジェクトの音データを合成する

5 まとめと今後の課題

本論文では動画内に存在するオブジェクトの動きをオブジェクト化することにより，効果音の合成を効率化するシステムの改良を行った．

現在のシステムは複数の物体が現れる動画を扱うことができないという問題点があり，特徴量のクラスタリングを行うなどして扱える動画の幅を広げる必要がある．今後は上記の問題点の改善のほか，既存の動画編集ソフトを参考にし，使いやすいインタフェースの構築を行うなどより動画制作を効率化できるよう改良を進めていきたい．

6 謝辞

本研究の一部は，JSPS 科研費 23500114 の助成を受けたものである．

参考文献

[1] M. Cardle, S. Brooks, Z. Bar-Joseph, and P. Robinson. Sound-by-numbers: motion-driven sound synthesis. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*, SCA '03, pp. 349–356, Aire-la-Ville, Switzerland, Switzerland, 2003. Eurographics Association.

[2] J. N. Chadwick and D. L. James. Animating fire with sound. *ACM Trans. Graph.*, 30(4):84:1–84:8, jul 2011.

[3] C. Zheng and D. L. James. Toward high-quality modal contact sound. In *ACM SIGGRAPH 2011 papers*, SIGGRAPH '11, pp. 38:1–38:12, New York, NY, USA, 2011. ACM.

[4] 鈴木 喜也, 岡部 誠, 尾内理紀夫. 無音動画に対する効果音貼り付けシステムの試作. In *DEIM 2012*, 2012.