

## 科学研究費助成事業 研究成果報告書

平成 26 年 6 月 9 日現在

機関番号：12612

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23500114

研究課題名(和文)無音の動画像ショットへの音貼り付け技術の研究

研究課題名(英文)Technology of Sound Effect Synthesis for Soundless Video

研究代表者

尾内 理紀夫(ONAI, RIKIO)

電気通信大学・情報理工学(系)研究科・教授

研究者番号：70323871

交付決定額(研究期間全体):(直接経費) 3,900,000円、(間接経費) 1,170,000円

研究成果の概要(和文):効果音を対象とする無音動画の有音化の研究を行った。効果音を、繰り返し音、瞬間的な音、持続性のある音の三種類に分類し、瞬間的な音と持続性のある音について研究した。瞬間的な音はミリ秒単位で発音位置を調整し、動画内の物体の動きと効果音の発音タイミングを一致させる必要がある。そこで音付き動画から効果音の合成に使用する特徴量と音データを切り出し、動画と独立なオブジェクト化を図ることとし、瞬間的な音の貼り付け技術を確立した。爆発音など数ミリ秒から5秒程度持続する音は、広周波数帯域にわたり不規則に音成分が現れる。このため無音動画内の物体の動きに合わせて持続性のある効果音を低劣化で伸長させる技術を確立した。

研究成果の概要(英文): We researched the technology of sound effect synthesis for soundless video. We classified sound effects into three types, that is, sound repeatedly, instantaneous sound, and persistent sound, and studied the persistent sound and instantaneous sound. In the case of instantaneous sound, it is necessary to match the timing in milliseconds of sound effects and movement of an object in the video. So from the video with sound, we cut out the sound data and the sound feature used for the synthesis, embed the cut out data in the objects independent of the video, and established technologies of instantaneous sound synthesis. In the case of sound that lasts about 5 seconds from a few milliseconds, such as explosions, the sound components appear irregularly over a wide frequency band. We established technologies of extending the persistent effect sounds in less degradation in accordance with the movement of an object in the video.

研究分野: 総合領域

科研費の分科・細目: 情報学 メディア情報学・データベース

キーワード: マルチメディア技術 音処理 効果音編集

### 1. 研究開始当初の背景

動画、特にアニメの製作現場においては、無音の動画ショットに適切な音を効率よく貼り付けていく作業が大きな負担となっており、これの軽減のための技術は重要であると考えられていた。一方、動画像と音(MIDIではない音)の双方の特徴点の対応を取ることを目指した技術の研究はなかった。また、音の貼り付けに際しては、画像内容だけではなく、画像内容と音との関係を解析し、双方の対応する特徴を抽出し、同期を取る必要があるが、そのような研究はなかった。動画像の編集システムは、Adobe Premiereをはじめ、既に商用システムがあった。しかし、例えば、人が歩いている動画像に、足音を貼り付ける際に、靴がフロアに当たるタイミングを取りながら、音の伸縮の際の音の高低変化をも制御しつつ、編集するシステムはなかった。すなわち、音を貼り付けるときに、音と画像内容の同期をコンピュータが支援しつつ貼り付けるシステムはなかった。

そこで、無音の動画像ショットに対して、画像にふさわしい音を、対象動画像ショットの内容との同期を取るなどの編集をし、適切な音を動画像ショットに貼り付けるという技術の確立を目指すという研究を着想した。

### 2. 研究の目的

無音動画像ショットへの適切な音貼り付け作業はきわめて時間のかかる作業である。そこで、音を無音動画像内容と合わせる同期機能、音を適切な長さに伸長する編集機能等の要素技術を確認し、作業負担を軽減することを研究の目的とした。

### 3. 研究の方法

無音の動画像ショットに適切な音を貼り付ける技術の確立という目的達成のため、以下のような方法、方針を採用した。

従来、音の編集においては、音声と音楽に関する研究は多数行われているが効果音に関する研究は少なかった。しかし効果音貼り付けは無音動画の有音化における有用性が高いため、音声、音楽以外の効果音を対象として研究を進めることとした。事前の22年度までの検討により、動画製作で使用される効果音を、繰り返し音、瞬間的な音、持続性のある音の三種類に分類していたが、風の音や車の走行音などの繰り返し音は本研究代表者、研究分担者らの先行研究によりその技術がすでに確立されていたため、残りの二種類の瞬間的な音、持続性のある音についての研究を主に展開することとした。

既存の方法を用いることが可能な部分は、極力、それを活用し、研究の軸である同期機能、編集機能に注力する。実際、システムの前処理部は既存技術を用いて構成した。

初年度である23年度はプロトタイプングを行い、24年度以降、効果音の種類を増加させるとともに、プロトタイプシステムの改良

を行うことにより、無音の動画像ショットに適切な音を貼り付ける技術を確認するという方法を探ることとした。

23年度では特徴量として各フレーム画像におけるオプティカルフローをまずは用いることとした。

音と動画の対応を取る編集手法としては、手法を固定的に考えるのではなく、各種の手法を検討することとした。検討の結果、物体の速度検出に、Kalalらによって提案されたTLD (Tracking Learning Detection) という手法を利用し、物体の出現・消滅の検出にはテンプレートマッチングを採用した。ボールのバウンド音、靴音などの音をデータベース化し、類似性などの判定にはソナグラムを使用することとした。

最終年度である25年度はシステムの評価に主眼を置くが、効果音の種類を増加させ、システムの有効性を向上させることとした。

### 4. 研究成果

(1) 23年度は、2種類のシステムのプロトタイプを作成した。どちらも効果音として、瞬間的な音を主たる対象とするものである。

一つ目のシステムは、効果音を鳴らすタイミングや音量・再生時間の調節を半自動的に行うことで、動画製作における作業時間を短縮する効果音付加支援システムである。ユーザは付加させたい効果音と動画を読み込ませる。次に音が鳴る源となる物体(例:バウンド音を付加させ

るならばボール)を選択する。物体を選択した後、検出したいタイミングとして、バウンド音のように物体の軌道・速度が変化したときに鳴る音の場合と、花火の音のように物体が出現もしくは消滅したときに鳴る音の場合のどちらかを選択する。その後、システムがタイミングを算出する。物体を選択しなければ、汽笛の音のように物体の動きに左右されずに鳴る音と判断し、ユーザが指定した区間にランダムに配置、もしくはその区間に合うように効果音の再生時間の調節を行う。なお、入力動画に既に音が入っていた場合、その音と効果音がどちらも聞こえるように付加を行う。タイミングを算出した後、検出できなかったタイミングの付加や余分な場所の削除、音量の調節を行い、効果音付加動画として出力する。物体の速度を検出するために、Kalalらによって提案されたTLD (Tracking Learning Detection) を物体追跡手法として使用した。

もう一つのシステムでは、類似した動きの物体は類似した音を発するという仮定に立ち、音の貼り付けを行う。システムは、まず既存の音付き動画と無音動画を読み込み、それぞれの動画から画像特徴量を抽出する。続いて、無音動画の画像特徴量と既存の音付き動画の画像特徴量とのマッチングをとり、類似した部分の対応付けを行う。マッチング結果を基に、音データを切り抜いて、無音動画に

貼り付ける。画像特徴量としてはオプティカルフローを採用した。ボールのバウンドの例では、音が適切な箇所に貼り付けられることを確認した。

(2) 24年度は、前年度までに開発した無音動画に対して効果音合成を行うシステムを改良し、実装、評価した。23年度に対象とした足音や物体の衝突音など瞬間的な音は、持続時間が一瞬しかない点が特徴である。そのため動画製作においてはミリ秒単位で発音位置を調整し、動画内の物体の動きと効果音の発音タイミングを一致させる必要がある。本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が、23年度までのシステム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点があった。そこで24年度は、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というアイデアを発想し、それを実現することにより、システムを改良した。本改良を実装した結果、より自然な音の合成が可能になるほか、効果音の差し替えや加工も可能となった。改良前後の比較評価実験を行い、より適切に効果音の合成が行われることを確認した。

具体的な改良点について次に述べる。

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。前年度までのシステムは、音付き動画から抽出したすべての特徴量を使用していたが、実際に効果音の合成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ることが24年度の冒頭の分析により明らかとなった。そこで、合成に使用する特徴量と音データを保持した、動画とは独立したオブジェクトを作成するというアイデアを発想した。また、前年度までは、マッチングにより選択された部分の音付き動画の音を断片化して貼り付けるため、合成する音をユーザ側がWavファイルに差し替えるといった操作ができなかった。24年度のオブジェクト化という改良により、動作における一連の特徴量と音データが1対1に対応する。これにより、音データのみを別のWavファイルなどに差し替える操作や音データを動作ごとに編集するといった操作が可能になった。

(3) 25年度は最終年度であり、前年度までのシステムを改良するとともに、研究をまとめた。効果音は、残された種類である持続性のある音に関して研究を行った。ガラスの破壊音や水はね音、爆発音など数ミリ秒から5秒程度持続する音は、広い周波数帯域にわたって不規則に音の成分が現れる点が特徴である。このため無音動画内の物体の動きに合わせて持続性のある効果音を劣化を抑

えて伸長させるアルゴリズムを考案した。本アルゴリズムは効果音の原音を受け取り、ユーザが伸長後の長さを入力として与えると次の手順で効果音の伸長を行う。まず、ユーザから与えられた伸長後の長さにより処理を分岐させる。この伸長率が2倍未満の伸長処理を行う場合は既存のタイムストレッチ系アルゴリズムによる伸長を行い終了する。そうでない場合、即ち、効果音を2倍以上に伸長する場合、本アルゴリズムは効果音の音量が最も大きくなる部分の前後(ピーク部分と呼ぶこととする)を原音から除去した波形を生成する。効果音の原音に、生成したピーク部分を除去した波形を重ね合わせることで効果音を伸長する。最後に伸長された波形に対し既存のタイムストレッチ系アルゴリズムを適用し、ユーザの入力した長さまで伸長を行い終了する。本アルゴリズムを実装し、その効果を確認した。同時に、システムとしてドラッグアンドドロップによる操作だけで効果音の伸長や発音位置の調整を行うユーザインタフェースも実装した。また音無しの楽器演奏動画に別の演奏音を貼り付ける検討も行った。

(4) まとめると、無音の動画像ショットへの効果音貼り付けに対して、本研究により確立された、音を無音動画像内容と合わせる同期機能、音を適切な長さに伸長する編集機能等の要素技術を用いることにより、既存の編集ソフトウェアと比較し、音の貼り付け作業に要する時間を大幅に短縮できることを確認した。結果、従来、無音動画像ショットへの適切な音貼り付け作業はきわめて時間のかかる作業であったが、その作業負担を軽減することを実現した。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表](計5件)

鈴木喜也、岡部誠、尾内理紀夫、無音動画に対する効果音貼付けシステムの開発、情報処理学会 第55回プログラミング・シンポジウム、2014年1月10日～12日、ラフォーレ倶楽部伊東温泉湯の庭(静岡県)

Tomohiro Yamamoto, Makoto Okabe, Yusuke Hijikata, Rikio Onai, Semi-Automatic Synthesis of Videos of Performers Appearing to Play User-Specified Music, 第21回 International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision 2014 (WSCG2014), 2013年6月24日～27日、チェコ共和国ピルゼン市

鈴木喜也、岡部誠、尾内理紀夫、無音動画に対する効果音貼付けシステムの検討、

日本ソフトウェア科学会 第 20 回インタラクティブシステムとソフトウェアに関するワークショップ WISS2012 デモ・ポスター発表、2012 年 12 月 6 日～8 日、古牧温泉青森屋(青森県)  
鈴木喜也、岡部誠、尾内理紀夫、無音動画に対する効果音貼付けシステムの試作、情報処理学会データベースシステム研究会、2012 年 3 月 4 日、シーサイドホテル舞子ピラ神戸(兵庫県)  
森下沙耶、岡部誠、尾内理紀夫、動画への効果音付加支援システムの作成、社団法人映像情報メディア学会メディア工学研究会、2012 年 2 月 18 日、関東学院大学(神奈川県)

## 6 . 研究組織

### (1)研究代表者

尾内 理紀夫 (ONAI Rikio)

国立大学法人電気通信大学・大学院情報理工学(系)研究科 教授

研究者番号：7 0 3 2 3 8 7 1

### (2)研究分担者

岡部 誠 (OKABE Makoto)

国立大学法人電気通信大学・大学院情報理工学(系)研究科 助教

研究者番号：4 0 5 5 7 2 1 1