社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

# 画像特徴と音声特徴の統合利用による結婚式映像のシーン分類

澤井　一基† 　高橋　友和†† 　出口　大輔† 　井手　一郎†,††† 　村瀬　洋†

† 名古屋大学 大学院情報科学研究科　〒 464-8601 愛知県名古屋市千種区不老町
†† 岐阜聖徳学園大学 経済情報学部　〒 500-8288 岐阜県岐阜市中鶉 1-38
††† 国立情報学研究所　〒 101-8430 東京都千代田区一ツ橋 2-1-2
E-mail: †{ksawai,ddeguchi,ide,murase}@murase.m.is.nagoya-u.ac.jp, ††ttakahashi@gifu.shotoku.ac.jp

**あらまし**　本報告では，画像情報と音声情報を利用して，結婚式映像中のシーンをイベントごとに分類する手法を提案する．近年，結婚式や結婚披露宴の様子を撮影した長時間映像における新郎新婦入場やケーキカットといった特定のイベント区間を効率的に閲覧する技術が求められている．しかし，特に結婚披露宴にはイベントの種類が多く存在し，多種多様な演出が含まれるため，分類が困難である．そこで，提案手法では複数の画像特徴と音声特徴を用いて，披露宴シナリオに記載された各種イベントと結婚披露宴映像の各部分とを対応付ける．これにより，結婚披露宴映像におけるシーン分類を行う．実験により，提案手法の有効性を確認した．
**キーワード**　結婚式，シーン分類，DP マッチング

# Scene Classification of Wedding Video Using Visual Features and Audio Features

Kazuki SAWAI†,

Tomokazu TAKAHASHI††, Daisuke DEGUCHI†, Ichiro IDE†,†††, and Hiroshi MURASE†

† Graduate School of Information Science, Nagoya University, Japan
†† Faculty of Economics and Information, Gifu Shotoku Gakuen University, Japan
††† National Institute of Informatics, Japan
E-mail: †{ksawai,ddeguchi,ide,murase}@murase.m.is.nagoya-u.ac.jp, ††ttakahashi@gifu.shotoku.ac.jp

**Abstract**　This report proposes a method for segmenting a wedding video into scenes according to events by using visual and audio features. Recently, there is a demand for a technique to efficiently browse specific event scenes such as bridal couple entering, cake-cutting in a wedding ceremony and party. However, it is difficult to classify wedding video scenes into events due to various styles and effects of the events especially in a wedding party. Therefore we classify a wedding party video into several events by matching the wedding party video to its scenario using visual and audio features. Experimental results demonstrate the effectiveness of the proposed method.
**Key words**　Wedding ceremony, Scene classification, Dynamic time warping

## 1. はじめに

結婚式や結婚披露宴は人生における特に重要なイベントの 1 つである．また，近年では撮影機器の急速な進歩により，結婚式や披露宴の様子をビデオカメラなどで撮影して長時間の映像として残すケースが増えてきている．それら映像中の特定のイベント区間をユーザが効率的に閲覧するためには，新郎新婦入場やケーキカットなど，結婚式や結婚披露宴のシナリオに記載された各種イベントと映像の各部分との対応付けが必要である（図 1）．そのため，結婚式・披露宴映像から特定のイベントに該当する部分を自動的に検出する技術が必要である．

一般に結婚式と呼ばれるものは，儀式としての結婚式と結婚披露宴から構成される．結婚式映像中のイベント検出に関する従来研究 [1] では，画像情報と音声情報を利用して，結婚式映像をそれらのイベントに対応した部分映像に分割している．映像を分割する際，画像情報を用いた処理としてカメラのフラッシュや花嫁衣装の検出，音声情報を用いた処理としてスピーチと音楽の分類 [2] を行っている．文献 [1] では，実際の結婚式映像を用いた実験において良好な結果が得られたと報告されているが，この研究はイベントの種類，順番が定型的な儀式とし

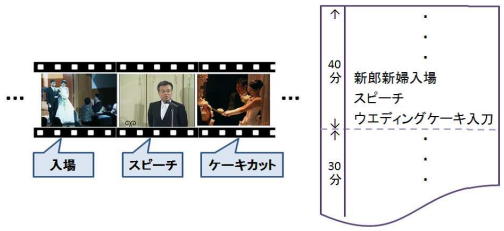


図 1　シナリオに基づく結婚披露宴映像のシーン分類

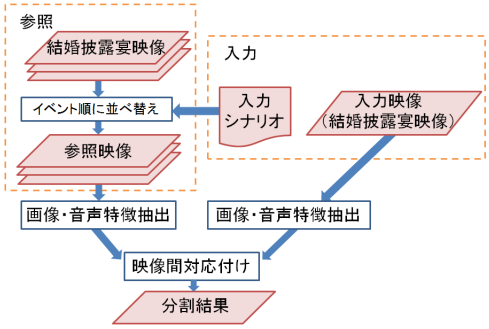


図 2　処理の流れ

ての結婚式のみを対象としている．そのため，イベントの種類が多く，多種多様な演出が含まれる結婚披露宴への適用は困難である．そこで，本研究では結婚披露宴映像をイベントごとに分割することを目的とする．

## 2. 提 案 手 法

提案手法では，複数の画像特徴，音声特徴を用いた入力映像とそれに付随する披露宴シナリオの対応付けにより，結婚披露宴映像のイベント分割を行う（図 2）．具体的には，他の結婚披露宴映像から作成した参照映像と入力映像との対応付けを行う．なお，参照映像は，事前にイベント分割を人手で行い，入力映像に付随する披露宴シナリオに沿ってイベントを並べ替えることによって作成する．また，参照映像と入力映像との対応付けに際しては，各映像を 10 秒単位の区間に分割し，各区間から画像・音声特徴を抽出し，区間単位で対応付けを行う．以下に提案手法で用いる画像特徴と音声特徴，映像間の対応付け方法について説明する．

### 2.1 画 像 特 徴

- 顔の数と位置：カスケード型識別器 [3] を用いて映像中の人物の顔を検出し，顔の数と位置を特徴量とする．
- フラッシュの頻度と強度：フラッシュの検出には輝度の時間変化を利用する．フラッシュ発生フレームは前後のフレームに比べて輝度が非常に高くなる．そこで，前後フレームとの輝度の差が閾値以上であればフラッシュ発生フレームとし，その発生頻度とフラッシュ強度（輝度の差が閾値以上の画素数）を特徴量とする．
- 輝度：フレームの平均輝度の平均と分散を特徴量として用いる．

### 2.2 音 声 特 徴

音声分類を行う従来研究 [2] で用いられている RMS（2 乗平均）と零交差の頻度を特徴量とする．また，従来研究 [2] による音声の分類結果（スピーチ・音楽・無音）も特徴量として用いる．

表 1　実 験 結 果

| 入力映像 | A | | B | | C | | 全体 |
|---|---|---|---|---|---|---|---|
| 参照映像 | B | C | A | C | A | B | |
| 一致率 [%] | 68.4 | 84.8 | 47.1 | 72.6 | 30.7 | 80.5 | 64.0 |

### 2.3 映像間対応付け

参照映像と入力映像の各区間から各特徴量を抽出し，DP マッチングにより，入力映像のイベント分割を行う．DP マッチングは各特徴を要素とする特徴ベクトル間のユークリッド 2 乗距離を用いて行う．

## 3. 実　　験

### 3.1 実 験 方 法

実際の結婚披露宴 3 回分の映像を用いて実験を行った．映像は 720 × 480 pixels，約 8 時間（計 2,832 区間），30 fps，MPEG 形式であった．入力映像と参照映像を変えながら計 6 回実験を行い，提案手法によるシーン分類精度を調べた．なお，参照映像は人手でイベント分割およびシーン分類を行った．

実験結果の評価には人手でシーン分類した結果を真値として，提案手法によるシーン分類結果との一致率を用いた．一致率の計算式を以下に示す．

$$\text{一致率 [\%]} = \frac{\text{正解区間数}}{\text{全区間数}} \times 100 \qquad (1)$$

### 3.2 実験結果・考察

実験結果を表 1 に示す．各入力に対する提案手法による一致率は平均で 64.0%であり，有効性が確認できた．

本実験では，特徴間のユークリッド 2 乗距離が小さい 2 つのイベントが短いイベントを挟んで存在するとき，誤対応がみられた．この誤対応は，その後のマッチングに影響を及ぼし，全体の一致率の低下要因となる．これに対しては，各イベントの長さを考慮し，DP マッチングに制約を与えることで改善できると考えられる．

## 4. お わ り に

本報告では，シナリオと入力映像の対応付けによる結婚披露宴映像のシーン分類手法を提案した．今後はイベントの長さを考慮した DP マッチングおよび複数の参照映像の利用による高精度なイベント分割手法を検討する．

**謝辞**　日頃より御討論頂く名古屋大学村瀬研諸氏および協力頂いた（株）クライムキューブに深謝する．本報告の一部は科研費による．本報告では MIST ライブラリ（`http://mist.murase.m.is.nagoya-u.ac.jp/`）を使用した．

## 文　　献

[1] W. Cheng et al., “Semantic-Event Based Analysis and Segmentation of Wedding Ceremony Videos,” Proc. MIR2007, pp.95–104, Mar. 2007.
[2] C. Panagiotakis and G. Tziritas, “A Speech/Music Discriminator Bsed on RMS and Zero-Crossings,” IEEE Trans. Multimedia, Vol.7, No.1, pp.155–166, Feb. 2004.
[3] P. Viola and M. Jones, “Rapid Object Detection Using a Boosted Cascade of Simple Features,” Proc. CVPR2001, Vol.1, pp.511–518, Dec. 2001.