

シナリオに基づく複数の事例映像との対応付けによる
結婚披露宴映像のシーン分割澤井 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 岐阜県岐阜市中鶯 1-38

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、結婚披露宴の式次第（シナリオ）に基づき、結婚披露宴映像をイベントごとのシーンに分割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て映像として残すことが増えている。特に結婚披露宴は結婚式に比べ長時間にわたるため、映像中の特定のイベント区間を効率的に閲覧するためには、式次第に記載された各種イベントを映像にラベル付けすることが必要である。そこで、提案手法では他の複数の結婚披露宴映像を式次第に沿って編集し直した事例映像との対応付けによって、式次第に記載されたイベントごとに結婚披露宴映像を分割する。実験により、提案手法の有効性を確認した。

キーワード 結婚披露宴, シーン分割, DP マッチング

Scene Segmentation of a Wedding Party Video
by Scenario-based Matching with Example VideosKazuki 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

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 party video into scenes of party events based on its program; a scenario. Recently, film in a wedding ceremony and its party has become common. Especially, because of the length, scenes in a wedding party video needs to be associated with events in the program for efficient browsing. Therefore we segment a wedding party video into several events by matching the wedding party video to example videos made by any wedding party videos under the program. Experimental results demonstrate the effectiveness of the proposed method.

Key words Wedding party, Scene segmentation, 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。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映像として残すことが増えつつある。特に結婚披露宴は結婚式に比べ長時間に渡るため、映像中の特定のイベント区間を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、式次第に記載された各種イベントと映像との対応付けが必要である。現在、この対応付けは式場の映像編集サービスの一貫として行われているが、

この作業は時間的にも、労力的にもコストがかかる。そのため、結婚披露宴のイベントと映像の対応付けの自動化に対する需要がある。これが自動化できれば、現在人手で行われている結婚披露宴のダイジェスト映像の作成などへの応用も期待できる。

結婚披露宴を撮影した映像は、図1(a)に示すように複数のシーンから構成される。

ここで、シーンとはある1つのイベントを撮影した映像区間を指す。本研究では、イベントと映像を対応付けることにより、結婚披露宴映像のシーン分割を行う。また、結婚披露宴は一般に図1(b)のようなイベントの順番を記載した式次第に沿って

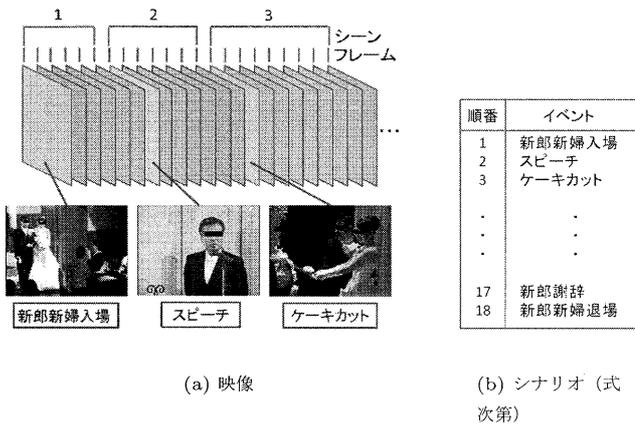


図1 結婚披露宴映像およびシナリオの構成

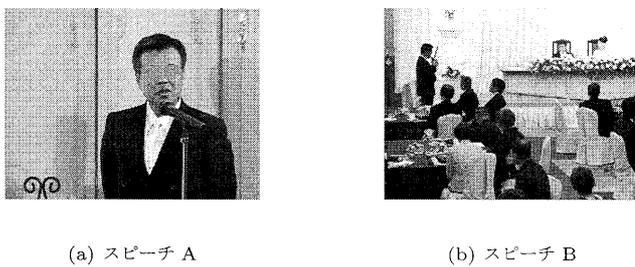


図2 異なる結婚披露宴における同じ種類のイベントの比較

進行される。本研究ではこのイベントの順番をシナリオと呼ぶ。結婚披露宴は以下のような特徴を持つ。

- 同じ種類のイベントであっても様々な演出がなされる
- 同じ種類のイベントであっても長さが異なる
- 披露宴によってイベントの順番が様々である
- 類似する特徴を持つイベントが複数存在する場合がある

図2は異なる2つの結婚披露宴における「スピーチ」シーン中の画像である。2つの画像を見比べると、同じ種類のイベントでありながら、構図や照明状況、背景や映っている人など様々な点が異なることがわかる。そのため、結婚披露宴映像のシーン分割においては、各イベントに固有の特徴を表すのに適した特徴量を選択的に用いる必要がある。また、同じ種類のイベントであっても披露宴によってシーンの長さが異なるため、このことを考慮した対応付けが必要である。一方、シナリオに記載された順番を利用することで、図3に示すように同じ披露宴での異なる種類のイベントが同じような特徴を持つ場合でも、シーン分割精度の向上が期待できる。このように、結婚披露宴イベントの特徴を考慮することが、高精度なシーン分割を実現する上で重要な要素となる。

本研究では、結婚披露宴映像の高精度なシーン分割を目的とする。手順としては、まず、シーン分割済みの他の結婚披露宴映像を、入力される披露宴のシナリオ中のイベントの順番に従って組み替えたものを複数作成し、それらを事例映像とする。その後、事例映像と入力映像から複数の画像特徴、音声特徴を抽出する。これらの特徴量を統合的に利用することで、各イベ



(a) キャンドルサービス (b) 新郎謝辞

図3 同じ結婚披露宴における異なる種類のイベントの比較

ントに固有の特徴を表現する。そして、これらの特徴を用いて事例映像と入力映像を対応付けることにより、イベント名がラベル付けされたシーン分割結果を得る。このとき、シーン間の対応付けにDPマッチングを用いることで、シーンの時間伸縮を許容し、特徴量の時間変化を考慮した対応付けを行う。

2. 関連研究

結婚式映像中のイベント検出に関する研究 [1] では、画像特徴と音声特徴を利用して、結婚式映像をイベント毎のシーンに分割している。映像を分割する際、画像特徴としてカメラのフラッシュや花嫁の検出、音声特徴としてスピーチと音楽の分類 [2] を利用している。本研究で対象とする結婚披露宴においても、カメラのフラッシュ、音声はイベントに固有の性質を表す特徴である。そのため結婚披露宴映像のシーン分割においてもこれらの特徴が利用できると思われる。しかし、結婚披露宴では花嫁の衣裳が多様であり、式中に複数回変化する場合もあるため、花嫁の検出を利用することは難しい。また、文献 [1] では、特徴量の時間変化をイベントごとにモデル化し、イベントと映像の対応付けを行っている。しかし、結婚披露宴では披露宴によって各イベントが様々な演出で行われるため、単一のモデルで表現することが難しい。したがって、結婚披露宴映像のシーン分割には、このようなイベントのモデル化による方法は適さないと考えられる。

特徴系列の時間伸縮を許容し、特徴量の時間変化を考慮した映像間対応付け手法として、DPマッチング (図4) がよく用いられている [5] [6] [7]。そこで、本研究ではDPマッチングを用いてイベントの時間伸縮を考慮した対応付けを行う。以下、DPマッチングによる映像間の対応付けの手順について説明する。映像Xを表す特徴ベクトル群 $\{x_1, x_2, \dots, x_P\}$ と映像Yを表す特徴ベクトル群 $\{y_1, y_2, \dots, y_Q\}$ に対して、DPマッチングを行うとする。ここでは、フレーム番号対 (i, j) 間の相違度 $d(i, j)$ として、特徴ベクトル間の Euclid 距離を用いる。つまり、

$$d(i, j) = \|x_i - y_j\|^2 \tag{1}$$

である。まず、累積相違度 $D(1, 1) = d(1, 1)$ とし、漸化式

$$D(i, j) = \min \begin{cases} D(i-1, j) + \omega_1 \cdot d(i, j) \\ D(i-1, j-1) + \omega_2 \cdot d(i, j) \\ D(i, j-1) + \omega_3 \cdot d(i, j) \end{cases} \tag{2}$$

によって累積相違度行列を更新していく。ここで、 $\omega_1, \omega_2, \omega_3$ は

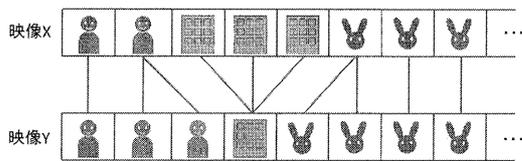


図4 DP マッチングによる映像間対応付け

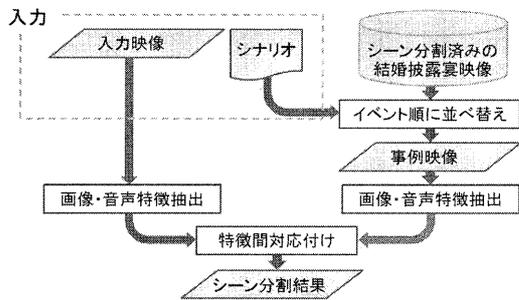


図5 処理の流れ

相違度に対する重みである。累積相違度行列を更新する際、選んだフレーム番号対 (i, j) を記録しておく。最終的な累積相違度 $D(P, Q)$ を2つの映像間の相違度として利用する。このとき、 $D(P, Q)$ を計算するまでに選んだフレーム番号対 (i, j) の系列が、 (x_1, y_1) から (x_P, y_Q) までの対応を表している。

結婚披露宴映像では、シナリオに基づいて事例映像を作成し、入力映像と事例映像に対してイベントに固有の特徴を利用したDP マッチングを適用することで、映像間におけるイベントの長さの違いを考慮した対応付けを行うことができると考えられる。

3. 提案手法

提案手法における処理の流れを図5に示す。手法は大きく以下の3つの処理からなる。

- (1) シナリオに基づく事例映像の作成
- (2) 画像・音声特徴抽出
- (3) 映像間対応付け

以降、処理の流れに沿って順に説明する。

3.1 シナリオに基づく事例映像の編集

事前の処理として、入力以外の複数の結婚披露宴映像をイベント毎に人手でシーン分割する。入力映像のシーン分割では、まず、入力以外のイベント名と入力となる結婚披露宴のイベント名の間で名寄せを行う。これにより、「ウェディングケーキ入刀」と「ケーキカット」などの同じ種類のイベントでイベント名が異なる場合に対処する。その後、シーン分割済みの結婚披露宴映像から入力シナリオに記載されたイベントを映したシーンを順に取り出す(図6)。本研究では、このようなシーンを入力シナリオに沿って編集しなおした映像を事例映像とする。提案手法では、事例映像と入力映像との対応付けによってラベル付きのシーンに分割する。

3.2 特徴抽出

提案手法では、複数の画像特徴・音声特徴を用いて入力映像と事例映像との対応付けを行う。図7に示すように T フレーム

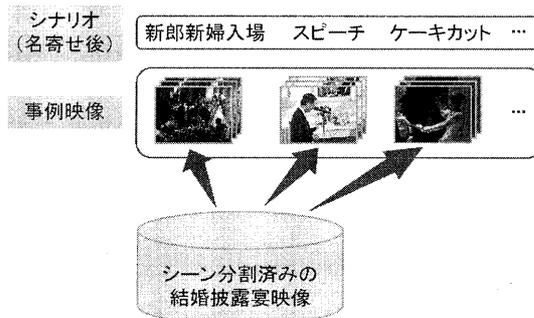


図6 事例映像の作成

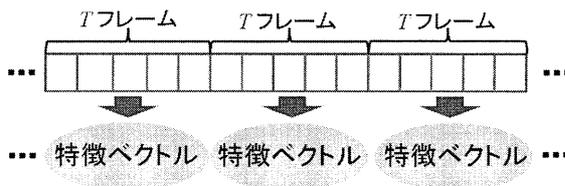


図7 特徴抽出

の単位区間から抽出した各特徴量を並べたものを特徴ベクトルとする。つまり、映像は(フレーム数/ T)個の特徴ベクトルで表される。

画像特徴として顔検出結果、カメラのフラッシュ検出結果、フレームの輝度を用いる。また、音声特徴として音声パワーの2乗平均(RMS)、零交差の頻度、音声のスピーチ/BGM/無音の分類結果を用いる。以下、各特徴の抽出手法について述べる。

(1) 顔検出

結婚披露宴においては、スピーチでは話者1人が大きく映されることが多く、ケーキカットでは新郎新婦が大きく映されるなど、顔の位置、顔の数がイベント固有の特徴を示す。そのため、提案手法では顔検出を行い、検出された顔の数、大きさの平均、分散、横方向の位置の分散、縦方向の位置の分散を特徴量として用いる。

顔検出には、顔検出器としてよく知られている Haar-like 特徴を用いたカスケード型識別器 [4] を用いる。

(2) カメラのフラッシュ検出

結婚披露宴では、新郎新婦入場やケーキカットなど、大きく盛り上がるイベントにおいてカメラ撮影が行われやすい。そのため、提案手法ではカメラのフラッシュを検出し、検出したフラッシュの数、フラッシュ発生前後フレームとの輝度差の最大値を特徴量として用いる。

カメラのフラッシュ検出には、フラッシュパターンに基づく映像照合に関する研究 [3] で用いられている手法を利用する。その手法の概要は以下のとおりである。

- (1) フレームごとにピクセルの平均輝度値を求める
- (2) フレーム間の輝度の変化を計算する
- (3) 変化がある閾値以上で、かつその前後フレームの平均輝度がほぼ等しい場合にフラッシュとして検出する

結婚披露宴は一般的に、室内の安定した照明状況下で行われるため、このような単純な手法で高精度にフラッシュを検出す

ることができる。

(3) フレーム輝度

結婚披露宴ではイベントによって暗転やスポットライトなど様々な照明効果が用いられる。また、新郎以外の男性は黒色のスーツを着用する場合がほとんどであり、イベントによっては彼らが画面内に占める割合が大きくなると考えられる。そのため、各フレームの平均輝度の平均、分散を特徴量として用いる。

(4) 音声特徴

結婚披露宴では、新郎新婦入場のときは大きな音楽が流れ、スピーチの際は人の話声が入るなど、イベントに固有の音声特徴を示す。そのため、音声のスピーチ/音楽/無音の分類を行い、その分類結果およびその結果を得るために用いた音声パワーの2乗平均 (RMS)、零交差の頻度を特徴量として用いる。

音声のスピーチ・音楽・無音の分類には、以下に示すような Panagiotakis らの手法 [2] を用いる。

(1) RMS に基づくセグメンテーション

まず、20ms を単位区間として、次式で表される音声パワーの2乗平均 (RMS) を求める。

$$\text{RMS} = \sqrt{\sum_{n=1}^N x^2(n)} \quad (3)$$

次に、1秒間の RMS のヒストグラムを次式で表される χ^2 分布で近似する。

$$p(x) = \frac{x^a e^{-bx}}{b^{a+1} \Gamma(a+1)}, x \geq 0 \quad (4)$$

ここで、パラメータ a, b は以下の式で表される RMS の平均 μ と分散 σ^2 から得られる値である。

$$a = \frac{\mu^2}{\sigma^2} - 1 \quad (5)$$

$$b = \frac{\sigma^2}{\mu} \quad (6)$$

隣接する1秒間の区間において式(4)で近似された分布間の距離が閾値以上であれば、異なる分類区間としてセグメンテーションを行う。

(2) RMS と零交差 (ZC) に基づく各セグメントの分類
まず、20ms を単位区間として、次式で表される零交差の頻度 (ZC) を求める。

$$\text{ZC} = \frac{1}{2} \sum_{n=2}^N |\text{sign}(x(n)) - \text{sign}(x(n-1))| \quad (7)$$

ここで、

$$\text{sign}(a) = \begin{cases} 1 & (a > 0) \\ 0 & (a = 0) \\ -1 & (a < 0) \end{cases} \quad (8)$$

である。その後、以下の評価基準を上から順に適用し、分類結果を決定する。

- セグメント内の RMS の中央値と平均値の和が閾値以下なら無音区間

- セグメント内の無音区間の割合が閾値以下なら音楽
- RMS と ZC を組み合わせた評価値が閾値以下ならスピーチ

- 周波数の最大値が閾値以上なら音楽
- RMS の分散が閾値以上ならスピーチ
- 上記のすべてに当てはまらなければ音楽

3.3 入力映像と事例映像の対応付け

3.3.1 相違度計算

映像間の相違度を、抽出した特徴ベクトル群間の DP マッチングにより計算する。2章では特徴ベクトル間の相違度として Euclid 距離を用いたが、結婚披露宴映像ではイベントごとに有効な特徴が異なると考えられるため、重みつき Euclid 距離を用い、その重みをイベントごとに変化させることで対応付けの高精度化を図る。つまり、2つの特徴ベクトル間の相違度 $d(\mathbf{f}_i, \mathbf{f}_j, \mathbf{w})$ を次式で計算する。

$$d(\mathbf{f}_i, \mathbf{f}_j, \mathbf{w}) = \sqrt{\sum_{k=1}^{K-1} w_k (f_{i,k} - f_{j,k})^2} + \begin{cases} 0 & (f_{i,K} = f_{j,K}) \\ w_K & (\text{other}) \end{cases} \quad (9)$$

ここで、 $K (= 13)$ は特徴ベクトルの次元である。また、音声の分類結果 $f_{i,K}$ については等しければ0、等しくなければ w_K として重みつき距離に足し合わせる。この相違度を用いて DP マッチングを行うことで、特徴ベクトル群 \mathbf{x}, \mathbf{y} 間の相違度 $D(\mathbf{x}, \mathbf{y}, \mathbf{w})$ を得る。

ここで、重み \mathbf{w} は事例映像からイベントごとに学習する。あるイベント E について学習したいとする。事例映像中のすべてのシーンの中で、イベント E を映したシーンの集合を $\mathbf{S}_E = \{\mathbf{S}_{E,1}, \mathbf{S}_{E,2}, \dots, \mathbf{S}_{E,N}\}$ とし、それ以外のシーン集合を $\mathbf{S}_O = \{\mathbf{S}_{O,1}, \mathbf{S}_{O,2}, \dots, \mathbf{S}_{O,M}\}$ とする。このとき以下の式で、イベント E についての重み \mathbf{w}_E を最適化する。

$$\arg \max_{\mathbf{w}_E} \left\{ \frac{1}{N C_2} \sum_{\mathbf{S}_{E,n}, \mathbf{S}_{E,m} \in \mathbf{S}_E} D(\mathbf{S}_{E,n}, \mathbf{S}_{E,m}, \mathbf{w}_E) - \frac{1}{N \times M} \sum_{n=1}^N \sum_{m=1}^M D(\mathbf{S}_{E,n}, \mathbf{S}_{O,m}, \mathbf{w}_E) \right\} \quad (10)$$

これにより、同じイベント間の相違度が小さくなり、異なるイベント間の相違度が大きくなるような重みをイベントごとに学習することができる。

3.3.2 特徴間対応付けによる映像分割

対応付けと分割は以下の3段階で行う。

(1) 初期位置の決定

イベント i の開始位置を p_{i-1} 、終了位置を p_i 、入力映像の始まりを P_b 、終わりを P_e とする。このとき、次の条件で式(12)を $\mathbf{p} = \{p_0, p_1, \dots, p_N\}$ について最小化する。

$$\text{条件} : p_0 = P_b, p_N = P_e, p_{i-1} < p_i \quad (11)$$

$$\arg \min_{\mathbf{p}} \left\{ \sum_{i=1}^N \frac{(p_i - p_{i-1} - \mu_i)^2}{\sigma_i^2} \right\} \quad (12)$$

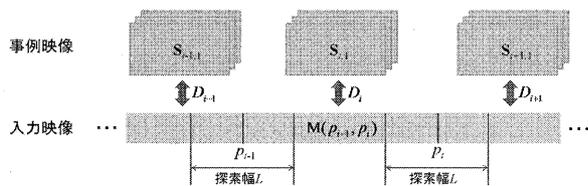


図 8 入力映像と事例映像の対応付け

ここで、 μ_i は事例映像中のイベント i の平均長さ、 σ_i^2 はイベント i の長さの分散である。式 (12) の最小化により、長さの分散が小さいイベントほど平均長さに近くなり、長さの分散が大きいイベントほどその長さが大きく伸縮するような位置 \mathbf{p}_0 を得る。最小化によって得られた位置 $\mathbf{p}_t(t=0)$ を初期位置とする。

(2) 事例映像との対応付けによる位置修正

対応付けの様子を図 8 に示す。事例映像中のイベント i を映した j 番目のシーンを \mathbf{S}_{i,j_i} とする。また、入力映像中の p_{i-1} から p_i までの区間を $\mathbf{M}(p_{i-1}, p_i)$ とする。このとき、 \mathbf{p}_t を初期位置として探索幅 L の範囲内で次式を $\mathbf{p}_{t+1} = \{p_0, p_1, \dots, p_N\}$, $\mathbf{j} = \{j_1, j_2, \dots, j_N\}$ について最小化することで、入力映像との相違度を最小にするイベント位置を探索する。

$$\arg \min_{\mathbf{p}_{t+1}} \min_{\mathbf{j}} \left\{ \sum_{i=1}^N \frac{(p_i - p_{i-1} - \mu_i)^2}{\sigma_i^2} + \sum_{i=1}^N \alpha_i D_i(\mathbf{S}_{i,j_i}, \mathbf{M}(p_{i-1}, p_i), \mathbf{w}_i) \right\} \quad (13)$$

ここで、 α_i は各イベントに対する重みであり、正解位置で式 (13) の値が最小になるように事例映像から学習する。式 (13) の第 1 項は式 (12) と同じものであり、長さの分散が大きいイベントほど大きく伸縮することを表している。第 2 項は入力映像に含まれる全イベントについての事例映像との相違度の重み付け和である。

(3) 繰り返し

(2) で得られた位置 \mathbf{p}_{t+1} と \mathbf{p}_t の差が閾値以上であれば、 t を 1 増やして、(2) に戻る。それ以外なら、 \mathbf{p}_{t+1} を最終的なシーン分割結果とする。

4. 実験および考察

提案手法の有効性を確認するために、実際の結婚披露宴映像を用いて 2 つの実験を行った。1 つ目の実験では、実際の結婚披露宴映像に対する提案手法のシーン分割精度を評価した。2 つ目の実験では、提案手法で用いた顔検出結果やフラッシュ検出結果など各特徴量について、その特徴量のみを用いてシーン分割を行い、各特徴量の有効性の検証を行った。以下、それぞれの実験について結果とそれに対する考察を述べる。

4.1 シーン分割精度

4.1.1 実験条件

異なる結婚披露宴を撮影した実際の映像 15 本を用いた。映像の仕様を表 1 に示す。また、全映像に含まれるすべてのイベント名とシーンの数を表 2 に示す。これらの映像に対し、シーン分割の単位区間長 T を 10sec, 対応付けの探索幅 L を 150sec

表 1 結婚披露宴映像の仕様

フレームレート	30 frame/sec
画素数	720 × 480 pixels
長さ	1.2 ~ 1.8 h
イベント数	11~ 19 個

表 2 全映像に含まれるイベント名とシーン数

イベント名	シーン数
新郎新婦入場	5
入場 (キャンドルサービス)	15
DVD 鑑賞	16
スピーチ	22
ケーキカット	5
お色直し退場	7
写真撮影	3
祝電披露	8
歓談	19
余興 (歌)	22
余興 (その他)	19
両親への手紙	15
花束贈呈	17
新郎謝辞	14
代表謝辞	14
新郎新婦退場	15
合計	216

として提案手法を用いてシーン分割を行った。各映像に含まれる区間数は平均で 542 区間、合計で 8129 区間であった。実験は 15 本のうち 1 本を入力映像として、それ以外の映像から事例映像を作成し、入力を変えながら 15 回行った。

4.1.2 評価基準

得られたシーン分割結果に対して以下の式で正解率を計算し、これに基づいて評価を行った。なお、真値は入力映像を人手でシーン分割したものとし、提案手法によるシーン分割結果が真値と一致した区間を正解区間とした。

$$\text{正解率} [\%] = \frac{\text{正解区間数}}{\text{入力映像の全区間数}} \times 100 \quad (14)$$

4.1.3 実験結果および考察

3.3.2 の (1) で決めた初期位置と提案手法による位置修正後のシーン分割精度を表 3 に示す。提案手法によるシーン分割の平均正解率は 82.6% であった。初期位置 (位置修正回数 0) での平均正解率は 69.8% であり、位置修正の繰り返しによる正解率の向上が見られた。正解率が最も高かったシーン分割結果の例を図 9 に示す。図 9 の色分けされた各区間は入力映像中の各シーンを表しており、初期位置と位置修正結果、位置修正結果と真値との対応する分割位置を点線で結んだ。初期位置に対して位置修正を行った結果、真値に近づいたことが分かる。このときの初期位置での正解率は 84.5% で、位置修正後の正解率は 95.7% であった。

同様に、正解率が最も低かった分割結果の例を図 10 に示す。このときの初期位置での正解率は 39.7%、位置修正後の正解率は 68.4% であった。正解率が低くなった原因としては、次の 2

表 3 初期位置と位置修正後のシーン分割精度

位置修正	あり (提案手法)	なし (初期位置)
正解率 [%]	82.6	69.8

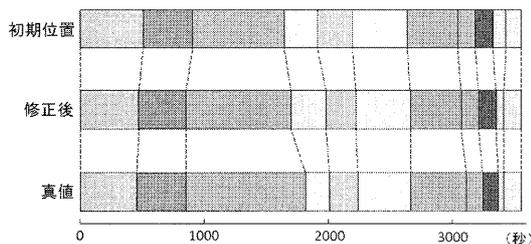


図 9 正解率が最も高かった例

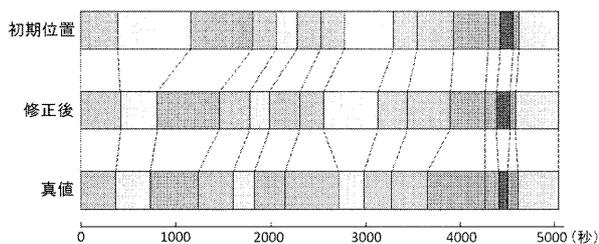


図 10 正解率が最も低かった例

点が考えられる。

- 同じ種類のイベントが連続する箇所が存在する
- 初期位置での正解率が低い

同じイベントが連続する場合、同じような特徴が続くため、特徴ベクトル間の相違度に基づく対応付けでは良い分割結果を得ることが難しい。このような場合には、司会者の音声検出や拍手検出などによるシーン分割境界の検出を行うことでシーン分割精度が向上できると考えられる。また、初期位置が真値から遠い場合、位置修正によって真値にたどり着く前に、局所的に相違度が小さい位置に誤って対応付けられてしまう可能性がある。これに対しては、式次第に記載された各イベントの大きな時間情報を利用して初期位置の精度を向上させることで、シーン分割精度を向上できると考えられる。

4.2 特徴量の有効性

提案手法で用いた各特徴量の有効性を調査するために、顔検出結果、フラッシュ検出結果、輝度、音声に関する特徴量のみを用いたシーン分割実験を行った。用いた特徴量以外の実験条件および評価基準は 4.1 で行った実験と同じである。ただし、式 (10) による各特徴量に対する重みの学習は行わなかった。それぞれの特徴量のみを用いたシーン分割結果の平均正解率を図 11 に示す。すべての特徴量において、初期位置での平均正解率 69.8% を上回る結果となった。これにより、各特徴量の有効性が確認できた。また、すべての特徴量において、各特徴量の重み付け距離を用いた対応付けによる平均正解率 82.6% には及ばなかった。このことから、すべての特徴量を組み合わせて用いることの有効性が確認できた。

5. おわりに

本報告では、シナリオに基づく事例映像との対応付けによる

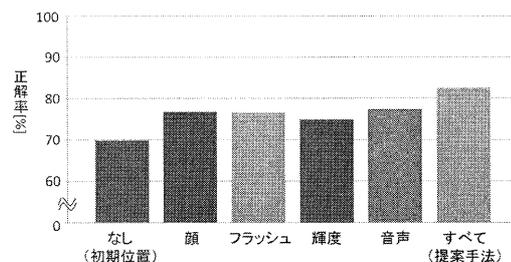


図 11 各特徴量を用いたシーン分割の平均正解率

結婚披露宴映像のシーン分割手法を提案した。実際の結婚披露宴を撮影した映像を用いて、提案手法によるシーン分割実験を行った。実験の結果、平均で 82.6% の正解率が得られた。また、提案手法で用いた各特徴量についてその有効性を確認し、それらの重み付け距離が有効であることも確認した。

今後の課題としては、同じイベントが連続して存在する場合のシーン分割精度の向上が挙げられる。これに対しては、シーン境界の検出を行うことで同じイベントが連続する場合であっても高精度なシーン分割ができると考えられる。また、式次第に記載されている大まかな時間情報を利用してシーン分割精度の向上が見込める。

謝辞 日頃より御討論頂く名古屋大学村瀬研諸氏に深謝する。本研究の一部は (株) クライムキューブのご協力及び科研費による。本研究では MIST ライブラリ (<http://mist.murase.m.is.nagoya-u.ac.jp/>) を使用した。

文 献

- [1] Wen-Huang Cheng, Yung-Yu Chuang, and Min-Chun Tien, "Semantic-Event Based Analysis and Segmentation of Wedding Ceremony Videos," Proc. International Workshop on Multimedia Information Retrieval 2007, pp.95-104, March 2007.
- [2] Costas Panagiotakis and Georgios Tziritas, "A Speech/Music Discriminator Based on RMS and Zero-Crossings," IEEE Trans. Multimedia, Vol.7, No.1, pp.155-166, February 2004.
- [3] 瀧本政雄, 佐藤真一, 坂内正夫, "フラッシュパターンに基づく映像照合手法," 第二回情報科学技術フォーラム (FIT 2003) 論文集, pp.177-179, September 2003.
- [4] Paul Viola and Michael Jones, "Robust Real-Time Face Detection," International Journal of Computer Vision, Vol.57, No.2, pp.137-154, May 2004.
- [5] 佐藤准嗣, 高橋友和, 井手一郎, 村瀬洋, "GPS 座標付き全方位映像群からの市街地映像マップの構築と街並変化の検出," 電子情報通信学会論文誌, Vol.J90-D, No.4, pp.1085-1095, April 2007.
- [6] Hiroyuki Uchiyama, Daisuke Deguchi, Tomokazu Takahashi, Ichiro Ide, and Hiroshi Murase, "Ego-localization using Streetscape Image Sequences from In-vehicle Camerasm," 2009 IEEE Intelligent Vehicles Symposium, pp.185-190, June 2009.
- [7] Chih-Yi Chiu, Cheng-Hung Li, Hsiang-An Wang, Chu-Song Chen, and Lee-Feng Chien, "A Time Warping Based Approach for Video Copy Detection," Proc. 18th International Conference on Pattern Recognition, pp.228-231, September 2006.