

複数人物の表情認識による 観客の満足度の自動定量評価

B4 川村涼太

目次

- 本研究の背景
- 実装
 - 流れ
 - 手法
- 実験
- 考察
- まとめ

背景

観客の満足度 → アンケート(従来)

↓ コストと手間がかかる

ビデオ映像から自動的に行われることが理想



顔検出や表情認識の手法は数多あるが、
多数の観客を対象としたものは当時存在せず



映像から観客の顔の有無と表情をパターン認識で認識させ、推定

実装(流れ)

観客の顔を検出する



表情を識別する



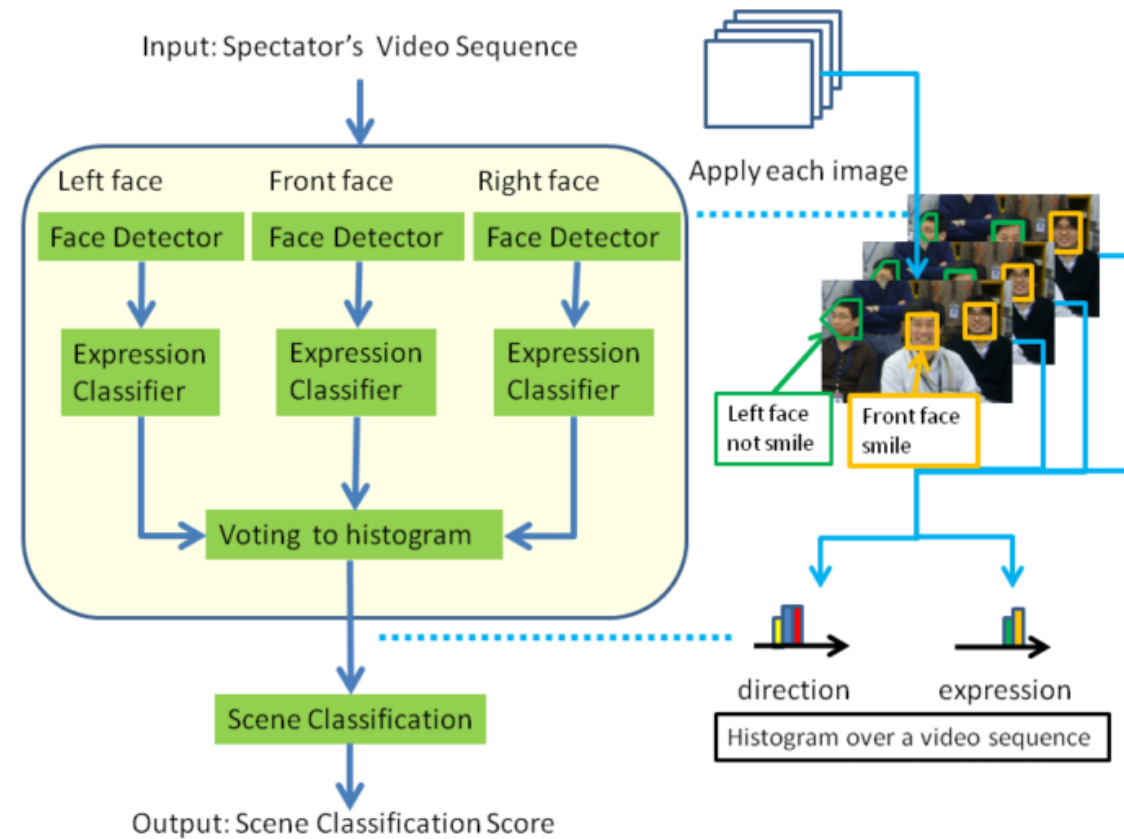
ヒストグラムに投票する



満足度ごとに分類する



スコアを算出する



手法

- ①顔検出器で動画の中から複数の聴衆の顔を検出
- ②表情識別器で顔の表情を認識
- ③満足度を推定

手法

- ①顔検出器で動画の中から複数の聴衆の顔を検出
- ②表情識別器で顔の表情を認識
- ③満足度を推定

①顔検出について

- 3種類の方向(左向き、正面、右向き)から検出
- 顔の向きと表情の頻度からヒストグラム特徴を生成
 - 画像パッチの量子化に相当
 - bag-of-visual-words法に基づいていると言える
- カスケード接続による顔検出器(VJ-detector)とSVMに基づく非顔フィルターの組み合わせ
 - 誤検出の確率を下げる

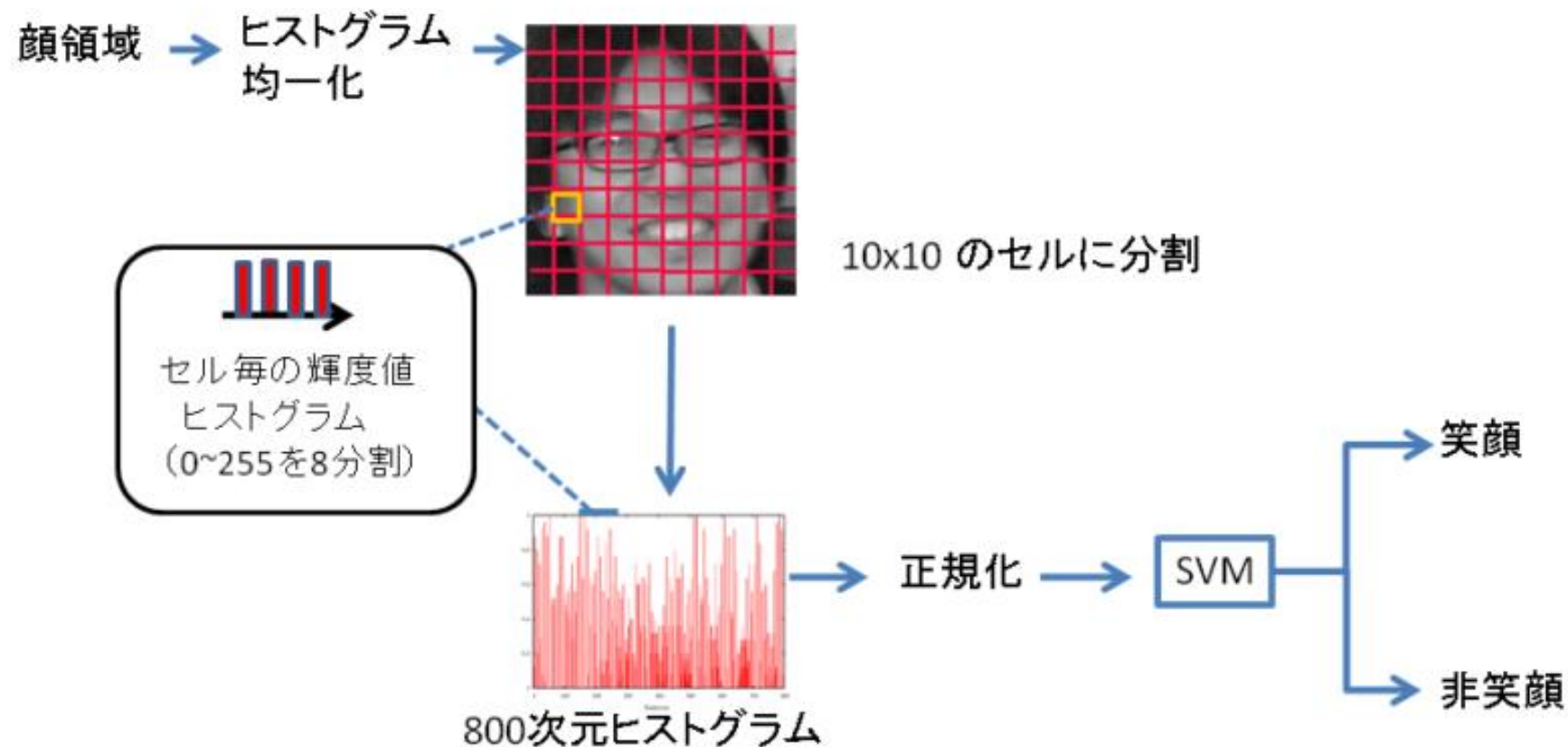
手法

- ①顔検出器で動画の中から複数の聴衆の顔を検出
- ②表情識別器で顔の表情を認識
- ③満足度を推定

②表情認識について

- 機械学習を使用
 - SVM(サポートベクターマシン)という学習モデルを使用
 - 訓練データ: 観客が「喜んでいるか否か」 and 「真剣に観戦しているか否か」を手動で画像シーケンスにラベル付けしたものを用意
 - 各フレームには3～30人の観客が映っている
- 画像の特徴量抽出には輝度ヒストグラム特徴(HI)を使用
 - HI特徴は単純な特徴ではあるが、次元数は800あり、表情認識には十分
 - $800 = 10 * 10$ (画像を重複の無い10*10のセルに分割) * 8(輝度値は0～255を8諧調に分割)

②表情の識別について



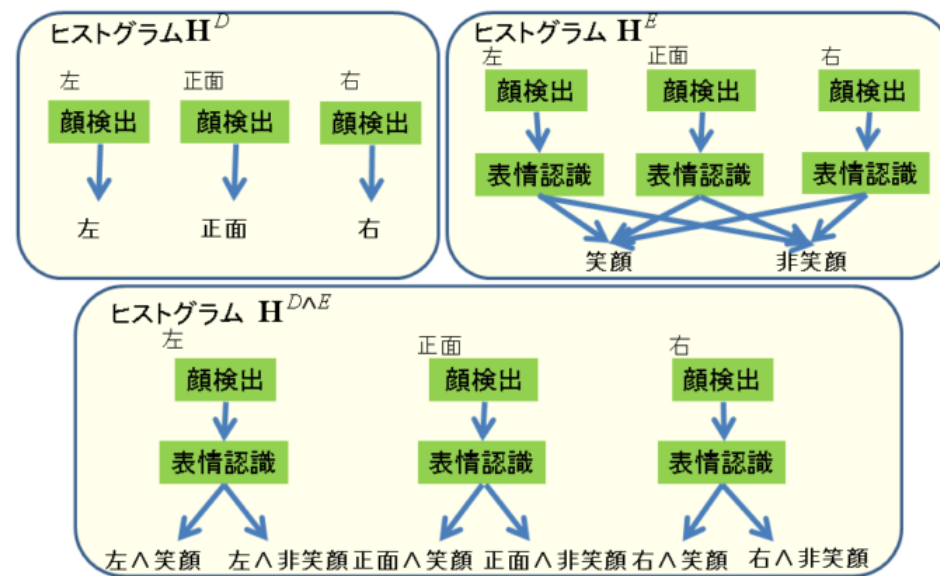
手法

- ①顔検出器で動画の中から複数の聴衆の顔を検出
- ②表情識別器で顔の表情を認識
- ③満足度を推定

③満足度推定について

- ①②の認識結果を基に、ヒストグラムに投票する

- 用いるヒストグラムは3種類
 - H^D (左向き、正面、右向きの3次元)
 - H^E (笑顔、非笑顔の2次元)
 - $H^{D \wedge E}$ (左向き \wedge 笑顔, 左向き \wedge 非笑顔, 正面 \wedge 笑顔, 正面 \wedge 非笑顔, 右向き \wedge 笑顔, 右向き \wedge 非笑顔の6次元)



③満足度推定について

- SVMで各ヒストグラムの重みを調整しながら算出



結果

【形】四角形：正面顔，五角形：横顔

【色】**橙色**：笑顔，**黄色**：非笑顔

実験

TVから得られた映像で実験した結果、

”喜んでいるか否か”

”真剣に観戦しているか否か”

をある程度識別し、そこから満足度を推定した

喜んでいるか否か→ 顔表情

真剣に観戦しているか否か→ 顔の向き

考察

- 横顔の表情認識率は、正面顔のそれより低かった
 - 横顔の多様性が理由
 - 誤認識の主な理由は、腕などによる隠れ領域
- 3チャンネルヒストグラムは、2チャンネルヒストグラム以上の識別率だった
 - 騒いでいる状況での精度が低かったので、新たな情報を追加して精度の向上を図る必要がある
- 喜んでいるが笑顔の少ない人が多いデータでは、満足度が低く推定された
- ミスラベルの混入により真剣度の識別率はやや低かった

まとめ

- bag-of-visual-wordsに基づいた顔識別器で表情を識別
 - 表情と顔の向きを投票したヒストグラムを作成
 - このヒストグラムを特徴ベクトルとして、満足度を自動推定
 - 今後は、顔表情と向きのクラスを増し、
 - より多くの観客を識別すること
 - より適切な観客のラベル付け(教師無し学習での分類など)
- が課題である