

写真スライドショー作成のための不要画像の自動検出

寛 弾[†] 岩野 公司[†]

東京都市大学[†]

1. はじめに

飲食店や小売店では、商品やサービスだけでなく、「場の雰囲気」も、顧客に快適な感覚を引き出すための重要な要素となる[1]。場の雰囲気を演出するための一つの手段として、店の雰囲気に合った写真画像をスライドショーとしてプロジェクタで投影する方法が考えられる。このようなスライドショーを作成する場合に、作成者がスマートフォンなどで日常的に撮影・収集した大量画像の中から、雰囲気に合った画像を選び出す作業が発生する可能性がある。しかし、日常的に撮影した画像の中には、一時的なメモとして撮影したものなど、スライドショーの素材として不向きな「不要画像」も含まれる。画像の選択の際には、このような不要画像を自動的に取り除いておくことが望まれる。

そこで本研究では、カフェなどの飲食店で使用するスライドショーの作成を想定し、大量の写真画像の中から「不要画像」を自動的に検出する手法の提案を行い、その性能を評価する。

2. 不要画像検出手法の提案

不要画像は、1) 写真中のメインとなるコンテンツが「文字」や「顔のアップ」といった内容である、2) 写真中のメインとなるコンテンツが曖昧で分かりにくい、などの特徴を持つと考えられる。そこで提案手法では、一般物体認識を用いたコンテンツ分析を行って撮影対象物の意味情報を抽出し、それを特徴量化して不要画像であるか否かの判定を行う。コンテンツ分析には、既存の動画認識クラウドサービスを利用する。様々なサービスが存在するが、本研究では映像検索における意味メタデータ付与[2]で他サービスよりも有効性が高いことが報告されている Google Cloud Vision API [3] を利用する。

図1に提案手法の流れを示す。まず、入力の写真画像を Google Cloud Vision API によりコンテンツ分析し、「Sky 96%, Sea 93%, Beach 93%, ...」の

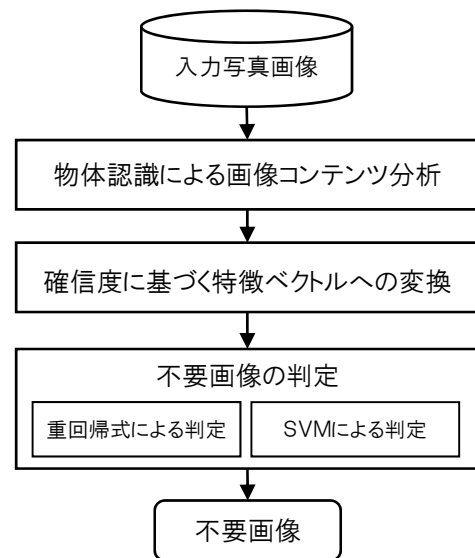


図1 提案する不要画像検出手法の流れ

ような、画像中の登場物の内容に関するキーワードとその確信度を得る。

次に、その情報を特徴量に変換する。コンテンツがどのような内容であるかを特徴量に反映させるため、まず、抽出されたキーワードを、「動物」「植物」「食物」「自然物」「風景」「人工物」「人物」「人体の部位」「文字」「芸術」「クリップアート」「抽象語」「その他」のカテゴリに分類し、「その他」を除く12のカテゴリについて確信度の最大値を求め、特徴量として利用する。また、撮影対象物が何であるかが曖昧な場合には、コンテンツ分析で得られる確信度が総合的に低くなると考えられるため、全キーワードに対する確信度の最大値と、12カテゴリの最大確信度の平均値を特徴量として利用する。これらの数値を用いて、入力画像を合計14次元の特徴ベクトルに変換する。

最後に、学習データで構築した識別器を利用して、不要画像か否かの判定を行う。今回は識別方法として、「重回帰式によって写真の相応しさを推定し、値が一定以下になるものを不要画像とする」方法と「SVM（Support vector machine）によって不要画像か否かを識別する」方法の2通りを試みた。



図2 実験に用いた画像の例（左：スライドショー使用画像として高評価なもの、右：低評価なもの）

3. 検出性能の評価

3.1 画像データの準備

評価実験のため、スマートフォンで撮影されたような写真として、風景や日常生活のシーン、画面のスクリーンショットなど、様々な内容の画像を40枚準備した。これらの画像に対し、大学生15名（男性11名、女性4名）によって、「カフェなどの飲食店で、良い雰囲気を演出するためのスライドショーで使用する写真として相応しいか」を5段階（1: どのような店にも相応しくない～5: どのような店にも相応しい）で評価してもらった。今回は、この平均評価値が2.5以下になる画像を不要画像と定義した。その結果40枚中の13枚が不要画像となった。

図2に、用いた画像のうちの平均評価値が高かったもの（左：平均評価値4.2）と平均評価値が低かったもの（右：平均評価値1.8）を示す。

3.2 評価実験の結果

性能評価は8分割の交差検証で実施した。

重回帰式による判定を行うため、「特徴ベクトルの各要素を説明変数として平均評価値を予測する重回帰式」を、学習データを重回帰分析することで求める。その際、不要な説明変数があると予測精度が落ちる可能性があることから、全40枚のデータに対する重回帰分析に変数減少法[4]を適用したところ、「動物」「自然物」「風景」「人体の部位」の4つの最大確信度が除外された。そこで、全ての数値を用いて14次元の特徴ベクトルを利用する場合の他に、変数削減後の数値を利用した10次元の特徴ベクトルを用いる場合についても性能評価を行った。

SVMは線形カーネルを用いる。コストパラメータは、学習データに対する8分割交差検証におけるグリッドサーチによって最適化を図った。ツールとしてはLIBSVM[5]を利用した。

表1に、重回帰式、SVMのそれぞれを使用した場合の不要画像検出の性能を、適合率と再現率で示す。結果を見ると、10次元の特徴ベクトルを用いたSVMによる判定を行った場合に最も

表1 不要画像検出の評価実験結果

識別手法	次元数	適合率	再現率
重回帰式	10	0.73	0.62
	14	0.67	0.62
SVM	10	0.71	0.77
	14	0.50	0.54



図3 不要と判定された画像の一例

総合的な性能が高くなった。このときの検出精度は82.5%（40枚中32枚が正解）であった。検出された不要画像の例を図3に示す。この例は、「文字」がメインのコンテンツであり、それにより不要画像と判断されたものと考えられる。一方で、文字で書かれているものであっても、その内容（意味）によっては人間が「要」と判断する場合もあるが、今回はそのような画像を不要画像として検出してしまう誤りも見られた。

4. まとめ

本研究では、飲食店などにおいて「場の雰囲気」を演出する写真スライドショーの作成を想定して、それに不要となる画像の自動検出手法の提案を行った。提案手法は、画像認識に基づくコンテンツ解析で得られた撮影対象物の意味情報を反映した10次元特徴ベクトルを用いて、SVMによって要/不要の判定を行うものである。実験の結果、判定精度は約8割となり、その有効性が示された。今後は、文字で書かれているものの内容を反映した特徴量の導入や、学習データの拡充による高精度化を検討する必要がある。

参考文献

- [1] P. Kotler, "Atmospherics as a Marketing Tool," *Journal of Retailing*, vol.49, no.4, pp. 48-64, 1973.
- [2] 板橋, 八木, "曖昧映像検索における意味メタデータ利用の検討," 映像情報メディア学会年次大会講演予稿集, vol.2018, 33B-1, 2018.
- [3] <https://cloud.google.com/vision/>
- [4] http://www.geisya.or.jp/~mwm48961/statistics/mul_reg3.htm
- [5] <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>