

バランスのとれた代表写真群選出のためのシステム

塩谷 祥加[†] 森下 奈保子[‡] 伊藤 貴之[†] 萩田 真理子[§]

^{†‡§}お茶の水女子大学 人間文化創成科学研究科 〒112-0012 東京都文京区2-1-1

E-mail: [†]{hiroka,itot}@itot.is.ocha.ac.jp, [‡]g1640602@edu.cc.ocha.ac.jp, [§]shagita@is.ocha.ac.jp

あらまし デジタルカメラやスマートフォン等の発達に伴い、誰もが気軽に多くの写真を撮影し保存することが可能となった。そこで写真データを整理したり、Web上で写真を共有したりする機会が増えてきている。一方で所有する写真データがあまりにも大量すぎて、整理や共有することに大変な労力を伴うこともある。そこで大量の写真群から代表的な写真群を適切に自動選出するシステムがあれば、写真の共有が手軽になると考えた。そこで我々は、多彩な特徴量を持つ写真を選出することにより、大量の写真群からバランスのとれた代表写真群を自動選出し提示する手法を提案している。我々の実装では代表写真群の選出結果を表示するユーザインタフェースも提供している。この画面上で対話的に「選出すべき写真」「選出すべきでない写真」を指定した上で代表写真群の選出を反復することで、ユーザの嗜好を反映して代表写真群の選出結果を調整することもできる。我々自身による先行研究では、各写真の特徴量を自動取得し評価値を算出した後、行列式を用いて代表写真群を選出していた。本報告では新たに、評価値を用いてグラフ彩色を応用することにより代表写真群を選出する手法を提案する。本報告では2種類の手法による結果を比較し議論する。

キーワード 写真選別、写真特徴量、ユーザインタフェース

1. はじめに

デジタルカメラやスマートフォンといった写真撮影機器の発達に伴い、誰もが気軽に多くの写真を撮影しては保存することが可能となった。しかし結果として大量の写真を保存することになり、それらを整理するには大変な労力が必要となる。さらにWeb上で写真を共有する機会も増加してきているが、大量の写真を全て共有するとしたら、アップロード時間とネットワーク負荷がかかる上に、サービスによっては容量制限などを伴う場合もあり困難を伴う。また大量の写真を全て公開しても閲覧者はその全てを閲覧するのに非常に多くの時間を要す。これらのことから、大量の写真の中から代表的な写真を選ぶ必要に迫られる機会が多い。しかし代表写真を選ぶ作業は、大量の写真を1枚ずつ目視確認して、好ましい写真とそうでない写真とに分類する非合理的な作業となる。そこで適切な枚数の代表写真群を自動選別するシステムがあれば、写真を楽しむユーザにとって写真の取捨選択が便利になると考えた。

我々は多彩な組み合わせの代表写真群を自動選出するための手法を提案している[1]。本手法では代表写真群の選出結果を表示するユーザインタフェースも提供している。この画面上で対話的に「選出すべき写真」「選出すべきでない写真」を指定した上で代表写真群の選出を反復することで、ユーザの嗜好を反映して代表写真群の選出結果を調整することもできる。

写真の自動選出手法に関して、従来の我々の手法では行列式にもとづく手法を採用していた。この手法では、まず各写真の特徴量を自動取得して評価値を算出し、著しく評価値が低い写真群を処理から省く。そして残りの写真群について、異なる特徴量を有する写真群をバランスよく選出する。特徴量のバランスのよさを「多次元空間にて特徴量ベクタによって形成される多面体の体積」と定義する。そしてユーザ指定の枚数に応じて仮の代表写真群を選出し、その写真群の特徴量ベクタが成す多面体の体積を計算する。この処理を反復した結果として体積が最大となる写真群の組み合わせを代表写真群とする。

本報告では新たに、グラフ彩色を応用した自動選出手法を提案する。この手法においては「どの写真ペアをとってきても類似度が低い写真群」がバランスのよい写真群であると定義する。はじめに自動算出した評価値を用いて、評価値合計

順に写真を並べ番号を振り、全ての写真のペアが張る面積を計算する。写真を点として、類似度が一定値以上のときにそのペアを辺で結びグラフを生成する。そしてグラフを彩色し、色1グループから写真番号が小さい順にユーザ指定の枚数を選出し代表写真群とする。

本報告ではこれら2種類の手法を用いた結果について比較し議論する。

本報告の構成は以下の通りである。2章では関連研究について述べる。続いて3章では提案手法を述べる。4章では本手法の実行結果について述べる。5章では本研究で得られた成果のまとめと今後の展望について述べる。

2. 関連研究

写真評価と写真の自動選出に関する研究は既に数多く発表されている。写真評価に関しては、写真の色や構図、ピントのぼけ具合といった一般的な特徴量をベースとした手法[2,3,4,5,6,7]がある。これらの手法においては、色が鮮やかな写真、被写体が中央に位置している写真、被写体にピントが合っている写真は評価が高い。また被写体となる各人物の表情や人数、人物間の写真上の位置などを評価要素として加えた手法[8,9,10]もあり、表情が良い、複数の人が写っている、人物どうしの距離が近く仲が良さそうな写真はより評価が高い。他には好ましい写真と好ましくない写真を分析することで評価基準を定め、写真を評価する手法[11]もある。

また写真評価結果にもとづいて写真に点数を付け、点数が高い順に写真を自動選出する研究[12,13]も発表されている。さらに、自動選出された写真をアルバムとして編集表示する商用サービス[14]もある。

3. 代表写真の自動選出手法

関連研究で紹介した写真評価手法に沿って評価が高い写真を選出すると、撮影状況によっては同じ人物や風景のみが選出されてしまう可能性が高い。そこで本手法においては、バランスのよい多彩な組み合わせの代表写真群を自動選出することを考える。具体的には、同じ被写体の写真が多数選出されることを防ぎ、さまざまな人物や風景、場面の写真からそれぞれバランスよい枚数で写真を選び出すことを目指す。

本手法では、写真群を大きく分けて以下の3つのグループに分類する。

第1群: 好ましい組み合わせの写真群

第2群: 第1群にも第3群にも含まれない写真群

第3群: 評価が低い写真群

N枚の写真を3グループに自動選別



図1: 写真群の自動選別

本手法ではまず、各写真について特徴量を自動算出する。この算出結果をもとにして各写真を評価し、評価が低い写真群を第3群に分類する。次に残りの写真群の中から、バランスよく多彩な写真群を選び出し第1群へ分類する。どちらにも分類されなかった写真を第2群に分類する。図1にその概念図を示す。

3.1. 写真の特徴量自動取得と評価値計算手法

バランスよくさまざまな写真を選ぶための基準の例として、我々は以下のような基準を考えた。

- ・ 写りの悪い写真（例えば暗すぎる、ピントが合っていないなど）は代表写真に選ばない。
- ・ 一人でも多くの人物が代表写真群の最低1枚に、いい表情をして写っているように代表写真を選ぶ。
- ・ 集合写真がある場合には、最低1枚の集合写真を代表写真に選ぶ。
- ・ 多くの場所を撮影している場合には、できるだけその各所の写真を代表写真に選ぶ。

これらの基準に沿って代表写真を選ぶために、現時点での我々の実装では、「人物の数」「人物の表情」「物体認識の情報」「ピント」「鮮やかさ」「撮影日時」の特徴量を自動で取得し、また「物体認識の情報」の特徴量を手動で算出する。これらの特徴量を用いて写真を評価し、0から1の値で点数を付けてファイルに保存する。また評価値が著しく低い写真群を第3群に自動分類する。

3.1.1. 人物の数

我々の実装では画像処理ライブラリOpenCVの顔検出機能を用いて矩形の顔領域を抽出し、その顔領域数によって人物の数を推定して特徴量とする。評価値の設定は以下の通りである。

- 検出された最大人数の80%以上：1
- 検出された最大人数の60%以上：0.8
- それ以下：人数×0.1

これにより、多くの人物の顔が写っている写真は優先的に代表写真として選出されやすくなる。

3.1.2. 人物の表情

我々の実装ではGoogleの画像認識APIであるCloud Vision[12]を用いることで表情の情報を取得している。特に笑顔認識の情報を採用し評価値の設定は下記の通りである。

- VELY_LIKELY：1
- LIKELY：0.8

UNLIKELY：0.4

VELY_UNLIKELY：0.2

取得不可：0

なお我々の実装では3.1.1項で判別した人物ごとに各写真における評価値を算出しその平均値を用いる。これにより人物が笑顔の写真が優先的に採用される。

3.1.3. 物体認識の情報

我々の実装では、Googleの画像認識APIであるCloud Vision[13]を用いることで物体認識の情報を取得している。取得した物体のラベル情報ごとに次元を作成し、各写真においてそのラベル情報が認識できた場合には”1”を認識できない場合は”0”とする。

3.1.4. ピント

我々の実装では、OpenCVを用いて人物の顔がどの程度はっきり写っているのかを計算する。検出された顔領域に対してエッジ検出を適用し、エッジと判断された画素の割合を計算する。ピントが合っていないほど値は小さくなる。ピント特徴量の評価値の算出は以下の通りである。

$$\begin{aligned} \text{ピント特徴量の評価値} \\ = \text{エッジ検出されたピクセル数} / \\ \text{顔領域のピクセル数} * 10 \end{aligned}$$

評価値が1以上の場合は”1”、0.1以下の場合は”0”とする。これにより、人物がぼやけていて鮮明でない写真は好ましくない写真と判定される。

3.1.5. 鮮やかさ

我々の実装では、各ピクセルのRGB値をHSV値に変換をする。そして彩度（S値）を写真全体での平均値を特徴量として求める。なお評価値は、平均値の1/100とし1以上の場合は”1”とする。これにより、色が鮮やかなほど、つまりSの平均値が高いほど評価が高く好ましい写真となる。

3.1.6. 撮影日時

我々の実装では、写真のexif情報を取得し、そこから撮影日時の情報を自動取得する。評価値はヒストグラムを用いることで算出する。ヒストグラムの区間幅は「(1番遅い時間-1番早い時間)/10」とし、度数が大きいほど重要な時間帯であるとすると、評価値は下記の通りである。

$$\text{評価値} = \text{度数} / \text{最大度数}$$

これにより、より多くの写真が撮影された時間の写真が優先的に選出される。

3.1.7. 人物の数

現時点では特徴量の取得を手動で行っているが、全ての写真から抽出される顔領域を人物ごとに自動判別できることを仮定している。認識できた人物ごとに次元を作成し、各写真においてその人物が認識された場合には”1”を認識されない場合は”0”とする。

3.2. 好ましい組み合わせ写真群の選出手法

本手法では、M次元の特徴量が与えられたN枚の写真の中から、バランスよい組み合わせのN'枚を選出することを想定する。バランスがよい組み合わせの写真群とはつまり以下の2条件

- ・ 評価が高い写真を多く選ぶ
- ・ 特徴が異なる写真を多く選ぶ

をできるだけ満たすように写真を選出することに相当する。3.2.1項では既に我々が提案している「行列式を用いた手法」[??]を紹介する。3.2.2項では本報告で新しく提案する「グラフ彩色を応用した手法」について論じる。

3.2.1. 行列式を用いた手法

1枚の写真を1本の特徴量ベクトルで表すと、特徴量ベクトルの長さは評価値、向きは特徴量の種類となる。図3は代表写真の選び方に関する概念を模式化したものである。図3(左)のように、短くてなす角も小さい特徴量ベクトル群に対応する写真群が選出された場合、評価が低くかつ似ている写真が多く選出されてしまう。逆に図3(右)のように、長くてなす角の大きい特徴量ベクトル群に対応する写真群が選出されれば、評価が高く多彩な写真が多く選ばれると考えられる。

図3(右)のような特徴量ベクトル群を選ぶための手段として本手法では、 N' 本の特徴量ベクトルで構成される多次元空間での多面体の体積を計算して、体積を最大にする特徴量ベクトル群を選出し、これに対応する写真群を代表写真群として第1群に分類する。

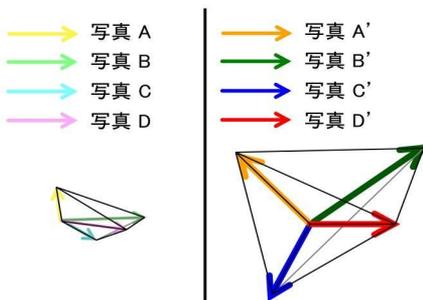


図3: 写真群の特徴量をベクトル化

N' 本の特徴量ベクトルで構成される多面体の体積を求めるにあたり、特徴量ベクトルが N' 次元であれば、 $N' \times N'$ 次元の行列を形成してその行列式を求めればよい。そこで $M > N'$ である場合には、特徴量ベクトルを M 次元から N' 次元に次元削減する。現時点での我々の実装では、代表写真の枚数 N' がユーザによって指定された時点で、以下の2段階処理を適用することで各写真の特徴量である M 次元ベクトルを N' 次元ベクトルに変換する。

- 1) まず、 M 次元ベクトルにおいて各次元ペア間でのPearson相関係数を求め、これが一定値以上となる特徴量ベクトルを1本に結合する。その結果、 M 次元ベクトルが N'' 次元ベクトル($M > N'' > N'$)に削減される。
- 2) 上述の N'' 次元ベクトルに対して主成分分析を適用し、上位 N' 個の主成分を採用することで、 N' 次元ベクトルに変換する。

続いて以下の手順によって何通りかの写真群に対して体積を計算する処理を反復し、体積が最大となる写真群を代表写真群として第1群に分類する。

1. ユーザが代表写真の枚数 N' を指定した時点で、全ての写真の特徴量を N' 次元ベクトルに変換する。
2. N' 枚の代表写真を仮に選出して $N' \times N'$ 次元の行列を構成し、掃き出し法によって行列式を算出することで体積を求める。
3. この処理を一定回数反復し、体積が最大となる代表写真群を正式に代表写真として第1群に分類する。

現在の我々の実装では、2.においてランダムに N' 枚の代表写

真を選出している。将来的には、ここに遺伝的アルゴリズムなどの進化計算手法を適用することで、限られた反復回数の中で最適に近い代表写真選出を実現できると考えている。

3.2.2. グラフ彩色を応用した写真選択

本報告では新しく、類似度の高い写真をエッジで連結することで生成されるグラフを用いて、代表写真を自動選択する手法を提案する。本手法では、写真ペアの非類似度は、2本の特徴量ベクトルがなす三角形の面積の大きさを表されるとする。言い換えると、三角形の面積が大きくなる写真ペアは差異の大きい特徴量を有していることから、類似していないと解釈する。そして本手法では、どの写真ペアをとっても類似度が低い写真群が、バランスが良い多彩な組み合わせの写真群であると解釈する。

以上の前提にもとづいて本手法では、以下の手順で代表写真を自動選出する。

- 1) 写真の特徴量ベクトルの長さでソートし、その順位を各写真に割り振る。我々の実装ではマンハッタン距離を採用しているため、特徴量ベクトルは単純に特徴量の絶対値の総和となる。
- 2) すべての写真ペアについて、2本の特徴量ベクトルがつくる三角形の面積を算出し、これを2写真間の非類似度とする。
- 3) 写真をノードとし、非類似度が閾値以下である写真ペアをエッジで連結することで、グラフを生成する。
- 4) グラフに彩色問題を適用する。結果として、類似する写真ペアに対応するノードには、異なる色が割り当てられる。
- 5) 同一色が割り当てられたノードの中から一定個数のノードを選ぶことで、多彩な特徴量を有する写真群を選択できる。我々の実装では単純に、1番目の色が割り当てられたノードを、特徴量ベクトルが長い順に選択する。

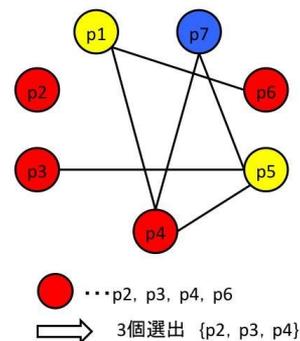


図4: グラフの彩色

図4は本手法による写真選択を図示したものである。写真群は特徴量ベクトルの長い順に p_1, p_2, \dots と番号づけられており、彩色問題を解くことで3色に色づけられている。そして赤色のノードの中から、番号の小さい順に (p_2, p_3, p_4) の3つのノードを選択し、それに対応する写真を代表写真とする。

3.3. ユーザによる手動選別

以上の処理によって全自動で写真を第1群から第3群に振り分けることも可能である。しかし、この結果に全てのユーザが満足するとは限らない。ゆえに、ユーザ自身による対話操

作で代表写真選出結果を修正できるように、写真を表示するユーザインタフェースを開発した。

図5にその概観を示す。この画面上でユーザは、移動する写真と固定する写真を選択することが可能である。例えば、代表写真に選出たくない写真が第1群に分類されている場合は、それを第2群に移動させることができる。また逆に、代表写真に選出したい写真が第2群に分類されている場合は、これを第1群に固定することが可能である。

ユーザが写真の移動操作を終えるたびに、「行列式を用いた手法」においては、3.2.1節で示した計算を再実行し、バランスよい組み合わせの代表写真群が再選出される。また「グラフ彩色を応用した手法」においては、ユーザが指定した写真と対応する点を1番目の色で彩色した後に残りの点を彩色する。そして1番目の色を割り当てられたノードの中から、ユーザが指定した写真を含むN枚が再選出される。

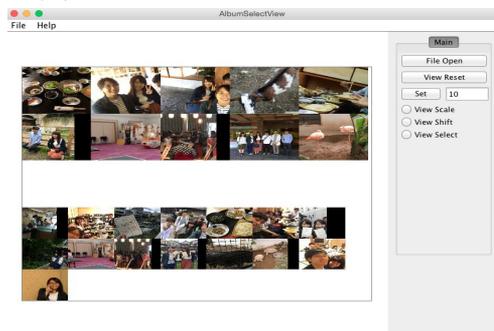


図5: 手動選択のためのユーザインタフェース

4. 実行結果

2種類の手法を用いて旅行写真25枚から5枚を選出した。用いた25枚の写真のうち2/3枚が人物メインの写真であり、それ以外は観光名所や食べ物等がメインの写真である。

行列式を利用した選出手法の場合、5枚の写真のうち集合写真が1枚、2人から4人の人物や観光名所が写った写真が3枚、食べ物の写真が1枚と比較的バランスがよい結果となった。しかし実行時間に45.8秒かかった。これは計算時間が以下のようになるからである。

・写真の選び方: $O(n^m)$

・掃き出し方による行列式計算: $O(m^3)$

グラフ彩色を応用した選出手法の場合、5枚全て参加者全員が写った写真であり、旅行を表すような特徴量を含んでおらず、バランスがいい写真群とはいえなかった。これは計算のはじめに評価値の合計を出しており、評価値が0か1で付いている人物認識の特徴量がそのまま反映されてしまったためであると思われる。一方で実行時間は5.9秒であり高速であった。これは計算時間が以下のようになるからである。

写真の選び方: $O(n)$

掃き出し方による行列式計算: $O(n^2)$

5. まとめと今後の課題

我々は大量写真群からバランスのとれた代表写真群を自動選出して提示する手法を開発している。本報告は自動選出のための2つ目の方法として、グラフ彩色を応用した手法を提案し、2種類の手法による自動選出結果を比較した。

本手法では写真を第1群から第3群に自動分類する。まず特徴量を自動算出することで写真を評価し、評価値が小さい写真を好ましくない写真として第3群に選別する。そして2種類の手法を利用して第1群の写真を選出する。我々は既に行列式を利用した手法[1]を提案しているが、この手法では指定された枚数の代表写真を仮に選択し、各写真の特徴量ベクトラ

によって構成される空間の体積を算出する。これが最大となる写真群が「最もバランスよくさまざまな写真を含む代表画像群」であるとして第1群に自動選別し、残りの写真群を第2群とする。本報告ではそれに加えて、グラフ彩色を応用した手法を提案しているが、この手法ではまず評価値合計順に写真を並べ番号を振り、全ての写真のペアが張る面積を計算する。写真を点とし、類似度が一定値以上のときにそのペアを辺で結びグラフを生成する。そしてグラフを彩色し、色1グループから写真番号が小さい順にユーザ指定の枚数を選出することで代表写真群とする。

さらに我々の実装では、ユーザの対話操作によって第1群と第2群の写真を入れ替えるユーザインタフェースを搭載している。対話操作結果に従って第1群に属する代表写真群を再度自動選出することにより、ユーザの嗜好を反映した代表写真選出を可能とする。

今後の課題として、人物認識と物体認識の情報に関する特徴量の評価値設定を改善したい。現状の実装では0もしくは1の2値で認識結果を記述している。しかし、この前提で特徴量ベクトラを生成すると、グラフ彩色にもとづく自動選出手法において、1となる変数が多い写真が選ばれやすい傾向が非常に強くなってしまふ。この点において、特徴量ベクトラの算出に改善の余地があると考えられる。また今回使用した写真とは異なる種類の写真群を用いることで、今回と同様な結果となるかを試したい。多様な写真群について本手法を適用して結果を検証することで、現状で認識している写真選出結果の問題点が正当であるのかを確かめたい。またユーザによる手動選別の回数を集計し、満足のいく写真選択結果が得られるまでにどの程度の手動選別が必要であるかを比較したい。現状で用いている写真群では枚数が少なく、ユーザによる写真選別はせいぜい1~2回程度であった。よって、さらに大量な写真群を用いた場合の手動選別の回数を検証したい。

参考文献

- [1] 塩谷祥加, 伊藤貴之, 萩田真理子, “バランスのとれた代表写真群を選出するためのシステム”, 第8回データ工学と情報マネジメントに関するフォーラム”, E5-5, 2016.
- [2] Ritendra Datta, Dhiraj Joshi, Jia Li, James Z. Wang, “Studying Aesthetics in Photographic Images Using a Computational Approach”, European Conference on Computer Vision, pp. 288-301, 2006.
- [3] Ritendra Datta, Jia Li, James Z. Wang, “Learning the Consensus on Visual Quality for Next-Generation Image Management”, ACM Multimedia Conference, pp. 533-536, 2007.
- [4] Yan Ke, Xiaoou Tang, Feng Jing, “The Design of High-Level Features for Photo Quality Assessment”, Conference on Computer Vision and Pattern Recognition, pp. 419-426, 2006.
- [5] Congcong Li and Tsuhan Chen, “Aesthetic Visual Quality Assessment of Paintings”, Sel. Top. In Sig. Proc., IEEE Journal of, vol. 3, no. 2, pp. 236-252, April 2009.
- [6] Yiwen Luo and Xiaoou Tang, “Photo and Video Quality Evaluation: Focusing on the Subject”, European Conference on Computer Vision, pp. 386-339, 2008.
- [7] Subhabrata Bhattacharya, Rahul Sukthankar, Mubarak Shah, “A Framework for Photo-Quality Assessment and Enhancement based on Visual Aesthetics”, ACM Multimedia International Conference, pp. 271-280, 2010.
- [8] Congcong Li, Andrew Gallagher, Tsuhan Chen, “Aesthetic quality assessment of consumer photos with faces”, Image Processing, pp. 3221-3224, 2010.
- [9] Shehroz S Khan and Daniel Vogel, “Evaluating visual

aesthetics in photographic portraiture”, Computational Aesthetics in Graphics, Visualization, and Imaging, pp. 1-8, 2012.

- [10] Shao-Fu Xue, Henry Tang, Dan Tretter, Qian Lin, Jan Allebach, “Feature Design for Aesthetic Inference on Photos with Faces”, Image Processing, pp. 2689-2693, 2013.
- [11] Yan Ke, Xiaoou Tang, Feng Jing School of Computer Science, Carnegie Mellon; Microsoft Research Asia, “The Design of High-Level Features for Photo Quality Assessment”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 419-426, 2006.
- [12] Congcong Li, Alexander C. Loui, Tsuhan Chen, “Towards Aesthetics: a Photo Quality Assessment and Photo Selection System”, Association for Computing Machinery’s annual conference on multimedia, pp. 827-830, 2010.
- [13] Lei Huang, Tian Xia, Ji Wan, Yongdong Zhang, Shouxun Lin, “Personalized portraits ranking”, ACM Multimedia Conference, pp. 1277-1280, 2011.
- [14] 富士フイルムYear Album,<http://year-album.jp>
- [15] GoogleCloudVision,<https://cloud.google.com/vision/>