

大量写真群からバランスのとれた代表写真群を選出するための一手法

塩谷 祥加[†] 伊藤 貴之[†] 萩田 真理子[‡]

[†]お茶の水女子大学 理学部 情報科学科 〒112-0012 東京都文京区 2-1-1

[‡]お茶の水女子大学 理学部 数学科 〒112-0012 東京都文京区 2-1-1

E-mail: [†] {hiroka, itot} @itot.is.ocha.ac.jp, [‡] hagita@is.ocha.ac.jp

あらまし SNS の普及に伴い、誰もが簡単に撮影した写真を共有することが可能となった。しかし大量の写真をすべて共有するのは、記憶容量やネットワーク負荷などの観点からも、また閲覧者が費やす時間の観点からも、必ずしも喜ばしいことではない。そこで多くの場面において我々は、大量の写真の中から適切な枚数の写真を選んで共有する。しかし、そのような代表写真群を適切に選ぶ工程は必ずしも簡単ではない。本報告では、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一手法を提案する。ここでいう「バランスのとれた」とは、値の異なる特徴量を持つ写真が同程度に選出されることで、多彩な組み合わせの代表写真を選出することを意味する。本手法ではまず各写真の特徴量を自動算出し、その結果から評価が低いと判断される写真群を処理から省く。続いて残った写真群から、ユーザ指定の枚数に応じて仮の代表写真群を選出し、特徴量のバランスのよさを判定する。この処理を反復した結果として、最も特徴量のバランスがよいと判定された写真群を代表写真群として自動選出する。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。

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

近年の SNS の普及により、誰もが写真を撮影しては Web 上で手軽に共有することが可能となった。もはや写真を共有することもコミュニケーションの一部と化している。食事の際や外出先での数枚から数十枚単位の少ない写真であればすぐに共有することができる。一方、旅行といった大きなイベントとなると数百枚から数千枚単位の写真を撮る人はたくさんいるだろう。

このような大量な写真の場合、その全てを Web 上で共有することは技術的困難を伴うことが多い。共有する枚数が多ければアップロードにもブラウザにも転送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る。さらに、大量の写真を全部公開しても閲覧者は全て閲覧するのに非常に大きな所要時間を要するし、重要な写真を探するのに時間がかかってしまう。そこで多くの場面において我々は、大量の写真の中から適切な枚数の写真を選んで共有する。しかし適切な枚数の写真を選ぶ際に、一般的には大量の写真を 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必要がある、非合理的な作業を必要とする。そこで我々は、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便利になる上に、コミュニケーションもより円滑になると考えた。

本報告では、バランスのとれた組み合わせの代表写

真群を大量の写真群から自動選出する手法を提案する。本手法ではまず各写真の特徴量を自動算出する。続いて特徴量から好ましい写真か好ましくない写真を評価し、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。好ましい組み合わせの写真群を自動選別する。この処理では、ユーザ指定の枚数に応じて仮の代表写真群を選出し、特徴量のバランスのよさを判定する。現段階の我々の実装では、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。この処理を反復した結果として、最も特徴量のバランスがよいと判定された写真群を代表写真群として自動選出する。また、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。

本稿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

写真評価に関する研究は既に数多く発表されており、一般的な特徴量をベースにした手法[1,2,3,4,5]と、被写体となる人物の写り方に着目した手法[6,7,8]に大別される。一般的な特徴量をベースにした手法では、写真の色や構図、ピントのボケ具合等に注目して写真を自動で評価している。色が鮮やかな写真、被写体が大きくもしくは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。人物の写り方に着目した手法では、被写体の人数、各人物の表情、人物間の写真上の位置などを評価要素として加えている。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。

また、写真評価結果にもとづいて評価が高い写真を自動選出する研究[9,10]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[11]もある。

3. 提案手法

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

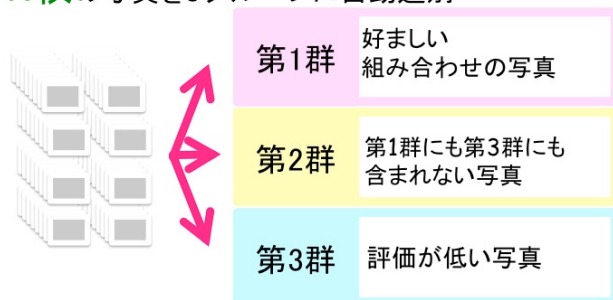


図 1: 写真群の自動選別

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の評価し、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」「ピント」「鮮やかさ」「撮影日時」の5種類の特徴量を算出する。これらの特徴量から写真を評価し、各評価値をファイルに保存する。また評価値が低い写真群を第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、これを特徴量とする。これにより、多くの人数の顔が写っている写真は優先的に代表写真に選出されやすくなる。なお我々の実装では、単に顔領域を抽出するだけでなく、全ての写真から抽出される顔領域を人物ごとに判別できることを仮定している。

3.1.2. 人物の表情

我々の実装では、detectFace();[12]という Web API を用いることで図2のような顔パーツの位置を取得している。

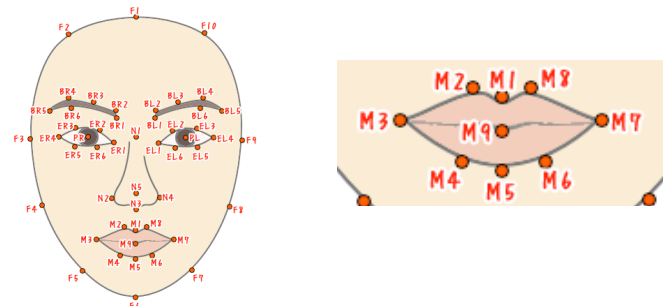


図 2: 顔パーツの位置

ここで、いい表情をしている顔は多くの場合にて、

口角が上がっていたり、口が開いていたたりすることから、M1 から M9 を用いて表情のよさを評価する。口の開き具合は M5 と M9 の y 座標の差で判断する。この差が大きければ口が開いていると判断できる。また口角の上がり具合は、M2 と M8 の y 座標値の平均と、M3 と M7 の y 座標値の平均の差で判断する。口角が上がっており、かつ口が開いている場合は最高の表情であるとし、口角が下がっている場合は好ましくない表情と判断する。

結論として、以下の 3 値

- ・ M5 と M9 の y 座標の差
- ・ M2 と M8 の y 座標値の平均
- ・ M3 と M7 の y 座標値の平均の差

を顔領域の大きさに応じて正規化して一次結合をとった値を、人物の表情に関する特徴量とする。

なお我々の実装では、3.1.1 項で判別した人物ごとに、各写真における特徴量を算出する。つまり写真群全体で K 人が写っていると判別されたら、人物の表情に関する特徴量だけで K 次元を有するものとする。なお、人物 A が写っていない写真 B において、人物 A に対応する次元の値はゼロであるとする。

3.1.3. 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。

$$\text{ピント特徴量} = \frac{\text{エッジ検出されたピクセル数}}{\text{顔領域のピクセル数}}$$

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。

3.1.4. 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。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。

3.2 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。ここでバランスよい組み合わせの写真を選ぶことは以下の 2 条件

- ・ 評価が高い (= 特徴量ベクトルが長い) 写真を多く

選ぶ

- ・ 特徴が異なる (= 特徴量ベクトルのなす角が大きい) 写真を多く選ぶ

をできるだけ満たすように代表写真群を選ぶことに相当する。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、評価が高くて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。

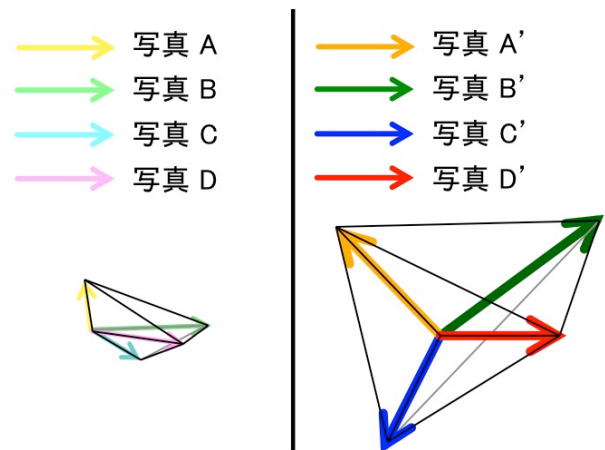


図 3: 写真群の特徴量をベクトル化

3.2.1. 特徴量の次元削減

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。

なお任意の写真群において必ずしも $M > N'$ であるとは限らない。特に写真群での登場人物数が少ない場合、あるいはユーザが非常に多くの写真を代表写真に選びたい場合、 $M < N'$ となることは容易に想定される。このような場合の処理については 3.3 節で後述する。

3.2.2.写真の仮選出と体積計算

現時点での我々の実装では、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。

3.3.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 4 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、代表写真に選出したくない写真が第 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。そして本手法は、ユーザが写真の移動操作を終えるたびに、3.2 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。



図 4: 手動選択のためのユーザインタフェース

なお、3.2 節で提案した手法について、 $M < N'$ であるときにはそのままでは実装できないことを論じた。これに対する解決策として、今後以下のような処理を実装する事が考えられる。

1. まずユーザインタフェース上の対話操作によってユーザが $(N' - M)$ 枚の写真を選択する。
2. 各写真の特徴量を M 次元ベクトルに変換し、 $M \times M$ 次元の行列を構成して M 枚の写真を自動選出する。

4.まとめと今後の課題

本報告では、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一手法を提案した。まず、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ル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。また本手法では、ユーザの対話操作に寄って第 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、特徴量取得の精度向上に取り組みたい。3.1 節に記述した特徴量ではまだ十分に写真を評価することはできない。また実装の精度が低いために、例えば人物の顔でないものを顔として認識してしまう場合がある。これらの問題点を解決し、さらに高い精度で写真を評価できるようにしたい。

また、ユーザの嗜好によって特徴量の重要度が異なることが想定されるため、特徴量の重要度をユーザごとに調節できるようにしたい。

さらに、操作時間と操作手順の削減に取り組みたい。現時点の実装では行列式の算出を反復することで無視

できない計算量が発生している．より少ない操作時間と操作手順でありながら，ユーザがより満足できる写真選出結果を得られるようにシステムを改善したい．

参 考 文 献

- [1] 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.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] Yiwen Luo and Xiaoou Tang, “Photo and Video Quality Evaluation: Focusing on the Subject”, European Conference on Computer Vision, pp. 386-339, 2008.
- [6] Congcong Li, Andrew Gallagher, Tsuhan Chen, “Aesthetic quality assessment of consumer photos with faces”, Image Processing, pp. 3221-3224, 2010.
- [7] Shehroz S Khan and Daniel Vogel, “Evaluating visual aesthetics in photographic portraiture”, Computational Aesthetics in Graphics, Visualization, and Imaging, pp. 1-8, 2012.
- [8] 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.
- [9] 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.
- [10] Lei Huang, Tian Xia, Ji Wan, Yongdong Zhang, Shouxun Lin, “Personalized portraits ranking”, ACM Multimedia Conference, pp. 1277-1280, 2011.
- [11] 富士フイルム Year Album, <http://year-album.jp>
- [12] detectFace(), <http://detectface.com>