

# ウェブ上の類似画像に付加されたメタ情報に基づく画像スコアリング

石原 のぞみ<sup>†</sup> 伊藤 雄一<sup>†</sup>  
高嶋 和毅<sup>†</sup> 岸野 文郎<sup>††</sup>

本論文では、大量の画像データベース内から魅力的な画像を検索するための基盤となる、人間の感性を考慮した新しい画像スコアリングを提案する。提案手法では、スコアリングの対象となる入力画像に対する類似画像を画像共有サイトから検索し、その類似画像に付加されたメタ情報（アクセス数、お気に入り数、コメント数、掲載日数など）をスコアリングに利用する。入力画像のスコアは、メタ情報から計算した各類似画像のスコアの、入力画像との類似度を重みとした加重平均として算出される。メタ情報は人間の主観評価と相関があると考えられるため、色や形などの低レベル特徴に基づく既存手法とは異なり、提案手法は人間の主観評価を考慮できる。評価実験により提案手法と参加者の主観評価の間に相関があることを示した。

## Attractiveness-based Image Scoring Using Similar Images on Websites

NOZOMI ISHIHARA,<sup>†</sup> YUICHI ITOH,<sup>†</sup> KAZUKI TAKASHIMA<sup>†</sup>  
and FUMIO KISHINO<sup>††</sup>

We propose a novel image scoring method that reflects human perceptions to retrieve desired images from large image databases. This method uses metadata attached to similar images from image sharing websites. Our method calculates the score of a new image (input image) as a weighted average of the scores of similar images from image sharing websites where each score of similar images is decided from metadata (e.g., the number of viewers, favorites, and comments), and each similar image score is weighted based on its similarity to the input image. Our method can reflect human perceptions or human subjective assessments unlike existing work that uses such low-level features as color and shape since metadata are expected to have correlations with human subjective assessments. The experiment showed a correlation between the scores by the proposed method and participant assessments.

### 1. はじめに

多くの人々は自分の画像データベースや、ウェブサイト上から魅力的な画像を検索したいと望んでいる。しかし、画像共有サイトやデジタルカメラの普及によりデータベース内の画像データ量が膨大になっているため、このような検索は困難である。Google 画像検索のような既存のテキストベースの画像検索や、色や形などの低レベル特徴に基づくコンテンツベースの画像検索では、人間の主観評価の高い魅力的な画像を効果的に検索できない。

魅力的な画像の検索を実現するためには、どのよう

な画像が魅力的だと判断されるかを知る必要がある。しかし、画像の魅力に関する基準は時代とともに変化するものであり、人間の主観も関係しているため、そのような基準を定義することは困難である。

画像の魅力を計算する手法（スコアリング）は今までにいくつか存在している。色や形といった低レベル特徴に基づく手法が多く存在するが<sup>(1),(2)</sup>、人間は画像の魅力を判断する際に、色や形だけでなく、コンセプトや構図なども考慮するため、このような既存手法は人間の主観評価を十分に考慮できない。

一方で、画像共有サイト上では、日々ユーザが手動で人間の主観に基づいたスコアを付与している。このスコアとはアクセス数、お気に入り数といったメタ情報であり、画像が魅力的であるほど、多くのユーザから視聴価値を認められ、そのアクセス数やお気に入り数は増える傾向にあると考えられる。こうして付与されるメタ情報は、ユーザの主観的な評価を反映していると言え、魅力的な画像であるかどうかをスコアリン

<sup>†</sup> 大阪大学 大学院情報科学研究科  
Graduate School of Information Science and Technology, Osaka University

<sup>††</sup> 関西学院大学 理工学部  
Science and Engineering Department, Kwansai Gakuin University

グする上で、有用な情報であると言える。こうした観点から、画像共有サイト Flickr は、独自の指標である *interestingness* をメタ情報から算出し、それに基づく画像検索を提供している。しかし、これはメタ情報が付与された画像のみが対象であり、所定のメタ情報を持たない画像をスコアリングできない。

そこで本論文では、人間の主観評価を考慮でき、さらにあらゆる画像をスコアリングできる新しいスコアリング手法を提案する。スコアリング対象の入力画像に対する類似画像を、画像共有サイトから検索し、検索時点でのその類似画像のメタ情報（アクセス数、お気に入り数、コメント数、掲載日数）を用いて入力画像をスコアリングする。これは入力画像の類似画像が魅力的であれば、入力画像も魅力的であると考えられるためである。この仮定に基づき、入力画像のスコアを、類似度を重みとして、メタ情報から計算された全類似画像のスコアの加重平均として計算する。提案手法は類似画像に付加されたメタ情報を利用しているため、メタ情報のない画像もスコアリングできる。

以下、関連研究、提案手法の詳細について述べ、評価実験により、提案手法によるスコアと実験参加者の主観評価に相関があるかどうかを評価する。さらに、提案手法によるスコアリングが写真家の写真の撮り方ガイドラインに沿っているかも検証する。

## 2. 関連研究

### 2.1 画像検索

画像検索は一般にテキストベース検索とコンテンツベース検索の二つに分けられる。

テキストベース検索では、クエリキーワードに関連するテキスト情報を持つ画像を検索する。Zhang らは、インターネット上のあらゆる画像フォーラムから画像を検索し、その周りのテキスト情報を取得し、その情報から検索クエリに近い画像を返す検索システムを提案している<sup>3)</sup>。この手法では、画像のタイトル、説明などをテキスト情報として利用し、さらに他の情報としてレーティングを利用している。

コンテンツベース検索では、クエリ画像の特徴量と類似する特徴量を持つ画像を検索する。特徴量としては色情報を利用する研究が存在する<sup>4)</sup>。また、Belongie は形状情報を特徴量として利用している<sup>5)</sup>。形状情報に基づく画像検索の精度は、画像の分割手法に大きく依存する。画像の分割手法としては標準カットが有名である<sup>6)</sup>。画像から特徴量を抽出した後は、その特徴

量を画像検索にどのように利用するかが問題となる。Ko らは領域レベルの画像分割と空間関係に基づくマルチ領域レベルの画像検索アルゴリズムを提案している<sup>7)</sup>。

コンテンツベース検索は、画像のコンセプトを考慮できないため、類似画像として検索された画像が、人間から見ると類似画像に見えない「semantic gap」と呼ばれる問題が生じる。そこで、この問題を解決するために、テキストベース検索とコンテンツベース検索を組み合わせた手法が多く研究されている。Chi らは、Google 画像検索のようなテキストベース検索のランキング結果を、コンテンツベースの特徴量も考慮した上で再ランキングする手法を提案している<sup>8)</sup>。コンテンツ要素としては、YCrCb カラーヒストグラムを利用している。

上に挙げた画像検索手法は、クエリ画像やクエリキーワードに関連のある画像のみを検索している。クエリに関連する画像を検索することはもちろん重要であるが、多くの人は魅力的な画像を好むため、魅力的な画像を検索する手法も重要である。

### 2.2 魅力に基づく画像スコアリング

魅力的な画像を検索するには、まずデータベース内の画像を魅力に基づいてスコアリングする必要がある。魅力に基づく画像スコアリングも従来から様々な研究がなされている。Savakis らは、画像の良さ、魅力について調べるため、複数の画像グループを用意し、参加者に各グループ内で画像をランク付けしてもらい、さらに、そのランク付けの理由を収集している<sup>9)</sup>。その結果に基づき、ポジティブな影響とネガティブな影響についてまとめている。これによると、影響の大多数は、人、構成、テーマといったものである一方、客観的な基準における影響に関してのみ、画質が画像の魅力と関連していると述べている。Wee らは画像の魅力として画像の鮮明さを使用し、画像の鮮明さを固有値によってモデル化している<sup>1)</sup>。Obrador らは、あらかじめ基準となる良い画像の構図を定義しておき、それに基づいて良い画像かどうかを判断してランキングする手法を提案している<sup>2)</sup>。良い画像の要素としては、色の調和、視覚的バランス、画質などを考慮している。また、Pedro らは人間の主観評価を考慮するために、画像共有サイト上のメタ情報（お気に入り数）を画像の魅力の指標として画像スコアリングする手法を提案している<sup>10)</sup>。この手法では、画像共有サイトから画像を収集し、低レベル特徴とテキスト情報を用いて Support Vector Machine でお気に入り数が閾値以上の画像とそうでない画像の特徴量を学習させる。この

学習モデルにより，入力画像が魅力的か魅力的でないかを分類することができる．また，回帰モデルにより低レベル特徴とテキスト情報からお気に入り数を計算する関数も求めている．この関数により，画像の魅力に基づくランキングを実現している．この手法は，お気に入り数との相関を実証しているが，人間の主観評価との相関を検証していない．

低レベル特徴に基づく既存手法は，人間の主観評価を考慮できないため，魅力に基づく画像スコアリングには不適切である．Pedro らは，画像共有サイト上のお気に入り数を魅力の指標とすることで人間の主観評価を考慮しようとしているが，人間の主観評価との相関を実証していない．さらに画像共有サイト上のお気に入り数以外のアクセス数やコメント数といったメタ情報も魅力に基づくスコアリングには有用であると考えられる．そこで，本論文では，人間の主観評価を考慮するために，画像共有サイト上の様々なメタ情報を利用した新しい画像スコアリング手法を提案する．提案手法は，類似画像のメタ情報を用いることで，メタ情報を持たない画像もスコアリングできる．

### 3. ウェブ上のメタ情報に基づくスコアリング

提案手法では，スコアリング対象の入力画像のスコアを計算するために，画像共有サイト上の類似画像のメタ情報を利用する．これは，メタ情報は画像共有サイトを利用するユーザによって付加された主観評価であり，主観評価を考慮した画像スコアリングに有用であると考えられるためである．さらに，類似画像のメタ情報を利用することで，メタ情報を持たない入力画像もスコアリングできる．類似画像検索では，入力画像の色，形などの画像特徴量と類似した特徴量を持つ画像が検索されるため，類似画像に付加されたメタ情報は，入力画像に付加されるであろうメタ情報と類似すると考えることができる．入力画像のスコアは，類似度を重みとして，メタ情報から計算された全類似画像のスコアの加重平均として計算する．本システムは，画像解析部，類似画像検索部，類似画像メタ情報問い合わせ部，スコアリング部から構成され，図 1 に示す手順で実行する．以降では各部の詳細について述べる．

#### 3.1 画像解析部・類似画像検索部

画像解析部では，類似画像を検索する前処理として入力画像の色，形などの低レベル特徴による特徴量を抽出する．抽出した特徴量を用いて，次の類似画像検索部で類似画像を画像共有サイト上から検索する．類似画像検索部では，入力画像の特徴量に対する，画像

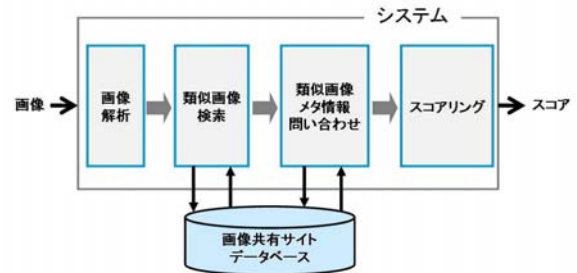


図 1 提案手法のフレームワーク

Fig. 1 Framework of the proposed method

共有サイト上の各画像の特徴量の類似度を計算し，類似度が高い画像を類似画像として返す．使用する画像共有サイトとしては，データベース内の画像量が十分であり，アクセス数，お気に入り数といった様々なメタ情報が各画像に付加されているものを使用する必要がある．そこで今回は，画像数が 50 億枚を超え，メタ情報も豊富な画像共有サイト Flickr を利用する．

画像解析部と類似画像検索部では今回は，既存の類似画像検索サイト Gazopa を利用する．Gazopa は入力画像の色，形などの特徴量に基づいて類似画像を検索するサイトである．検索された各画像には入力画像に対する類似度も付加される．また，検索結果はウェブ上のあらゆるサイトから検索した結果，動画の検索結果，Flickr から検索した結果などに分類できる．今回は，画像共有サイト上のメタ情報を使用するため，これらの検索結果のうち Flickr の検索結果のみを利用する．Gazopa では，入力画像に対して毎回 1000 枚の類似画像が検索され，Gazopa のデータベースが更新されない限り，返される類似画像は不変である．

検索結果にはスコアリングに不適切な画像も含まれている．例えば，入力画像との類似度があまりにも低い画像は類似画像とみなせない．ゆえに，類似画像としては，類似度が閾値以上の画像を使用する．この閾値を決めるために，被験者数人で，Gazopa によって検索された類似画像と入力画像を比較する予備実験を行った．その結果，類似度 40% 以下の類似画像は，主観的に見て入力画像と類似していると言えないことが分かり，今回は，検索結果のうち類似度が 40% 以上の画像を類似画像として取得することとする．また，アクセス数が 0 の画像は，誰にも評価されていないため，検索結果の中から除去する．この時，入力画像によって閾値以上の類似度を持つ類似画像の数に大きな差が生じる．特に Gazopa の検索結果は 1000 枚と豊富であるため，類似画像数の差が大きくなる恐れが

ある．提案手法では類似画像の数が豊富であるほどスコアリングの信頼性が上昇すると考えられるが，後の評価実験において画像間の比較を行う際の不平等が生じないために，参照する類似画像の数に上限を設けることとする．今回は類似度上位 100 枚のみをスコアリングに利用する．なお，類似度が下位であるほど，その画像のスコアは入力画像のスコアに反映されにくい本手法の特性から，上位のみを参照することで，閾値以上の全類似画像を参照した際のスコアに近いスコアを算出できると考えられる．

### 3.2 類似画像メタ情報問い合わせ部

類似画像メタ情報問い合わせ部では，Gazopa により取得した Flickr 上の各類似画像のメタ情報を取得する．今回，使用するメタ情報として，アクセス数，お気に入り数，コメント数，掲載日数，入力画像に対する類似度を取得する．

### 3.3 スコアリング部

スコアリング部では，取得した入力画像に対する類似画像のメタ情報を利用して，入力画像のスコアを計算する．まず各類似画像のスコアをメタ情報から計算し，類似度を重みとした全類似画像のスコアの加重平均として入力画像のスコアを最終的に計算する．ここで，類似画像のメタ情報を入力画像のスコアリングに使用するものは，入力画像の色，形などの特徴量と類似した特徴量を持つ類似画像に付加されたメタ情報は，入力画像に付加されるであろうメタ情報と類似すると考えられるためである．

#### 3.3.1 類似画像のスコア

メタ情報は人間の主観評価と相関があり，魅力に基づくスコアリングに有効であると考えられるため，類似画像のスコアは，付加されているメタ情報を利用して計算する．メタ情報を用いた式としては様々なものが考えられるが，どの式が最もスコアリングに適しているかわからない．そこで，今回は試験的に 10 個の候補式を考案し，実際にどの式が最も適しているかを評価実験で検証する．以下では，この候補式について述べる．

メタ情報のうち，お気に入り数やコメント数は，ユーザが画像を積極的に評価しようとした結果であるため，このメタ情報をアクティブメタ情報とする．まず，このアクティブメタ情報に着目した式を式 (1) から (8) に示す． $score1$  は画像を見た人のうちお気に入り登録している人が多ければ，それは良い画像とみなせることを考慮している． $favorites$  はお気に入り数， $views$  はアクセス数を表す．

$$score1 = \frac{favorites}{views} \quad (1)$$

次に， $score2$  では，お気に入り数がユーザの主観評価と大きく関係していると考え，お気に入り数をより重要視している．

$$score2 = \frac{favorites^2}{views} \quad (2)$$

しかし，上の二つの式は，お気に入り登録した人がその画像を何度も見る可能性を考慮していないという問題がある．そこで， $score3$  では，掲載期間が長くなればなるほどお気に入り登録した人はその画像を何度も見ると考え，その分お気に入り数の重要度を下げている． $days$  は掲載日数を表す．

$$score3 = \frac{favorites \cdot (1/days)}{views} \quad (3)$$

また，コメント数についてもお気に入り数と同じような考え方ができるため，

$$score4 = \frac{comments}{views} \quad (4)$$

$$score5 = \frac{comments^2}{views} \quad (5)$$

も考案できる． $comments$  はコメント数を表す．そして，お気に入り数とコメント数の両方を考慮し，それぞれに重みをつける式として，

$$score6 = \frac{\alpha \cdot favorites + \beta \cdot comments}{views} \quad (6)$$

$$score7 = \frac{\alpha \cdot favorites^2 + \beta \cdot comments^2}{views} \quad (7)$$

$$score8 = \frac{\alpha \cdot favorites^2 + \beta \cdot comments}{views} \quad (8)$$

を考案した． $score8$  は，コメント数よりお気に入り数のほうが主観評価と相関があると想定した式とする． $score6-8$  における  $\alpha, \beta$  は重みである．

以上のアクティブメタ情報を考慮した  $score1-8$  は，お気に入り数やコメント数が 0 である場合に， $score$  の値が 0 になってしまう問題がある．アクセス数は，ユーザが画像を積極的に評価しようとした結果ではないが，見るという行為により，暗黙的にその画像を評価した結果を表している．ゆえに，アクセス数を考慮することも重要であるため，上記の式にアクセス数を考慮した値を加える．加える値としては，ある掲載日数における平均アクセス数より，対象画像のアクセス数が多い場合は大きく，逆の場合は小さくなるような値とする．そこで，各掲載日数における平均アクセス数を求めるために，Flickr から画像を無作為に 10 万枚収集し，掲載日数とアクセス数のセットデータを収集した．収集したデータから，各掲載日数における平

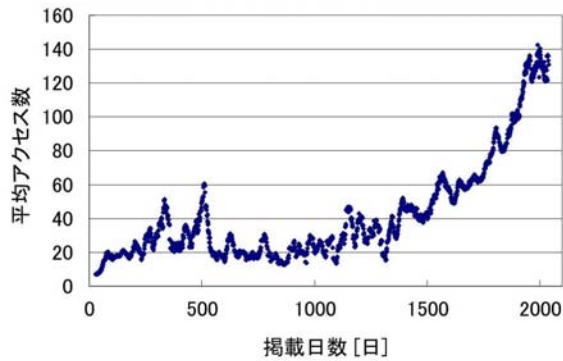


図 2 平均アクセス数の分布  
Fig. 2 Distribution of VIEW

平均アクセス数 (*VIEW*) を求め、散布図を求めた。平均アクセス数は正規分布に従うと仮定し、外れ値は除外した。散布図を平滑化するため、約 30 日間の移動平均をとった結果、図 2 に示す散布図が得られた。散布図は二次関数的な変化をしているため、掲載日数から平均アクセス数を計算する二次関数  $f$  を重回帰分析で求めた。

$$\begin{aligned} VIEW &= \\ &3.62 \times 10^{-5} days^2 - 1.84 \times 10^{-2} days + 21.0 \\ &= f(days) \quad (days > 0, R^2 = 0.86) \quad (9) \end{aligned}$$

この平均アクセス数を求める関数  $f$  を利用して、式 (10) のように  $score_{1-8}$  を補正する。

$$\begin{aligned} score' &= score + \gamma \cdot g(views, days) \quad (10) \\ g(views, days) &= views - f(days) \end{aligned}$$

以降では、補正して得られた  $score'_{1-8}$  を新たに  $score_{1-8}$  と定義する。関数  $g$  を加えることで、ある掲載日数における平均アクセス数より、対象画像のアクセス数が多い場合は  $score$  が大きく、逆の場合は小さくなる。ただし、 $g$  を単に加算するだけでは、 $g$  の値が大きすぎて、元の  $score$  の値が無視されてしまうため、重み  $\gamma$  を用いて元の  $score$  の桁に、 $g$  の桁を合わせている。

また、 $g$  はそのままでも  $score$  式となるため、 $score_9$  は式 (11) のように定義する。

$$score_9 = g(views, days) \quad (11)$$

最後に、1 日当たりのアクセス数が多いほど  $score$  を大きくするような式として、 $score_{10}$  を考案する。 $score_{10}$  を他の式と比較することで、アクティブメタ情報の効果を調べることができる。

$$score_{10} = \frac{views}{days} \quad (12)$$

以上の  $score_{1-10}$  を類似画像のスコアを計算する  $score$  候補式とする。

### 3.3.2 入力画像のスコア

入力画像のスコアは、類似度を重みとして、全類似画像のスコアの加重平均として計算する。加重平均は以下の式 (13) で表される。

$$\begin{aligned} SCORE &= \frac{1}{SS} \sum_{k=1}^n (sim_k \cdot score_k) \quad (13) \\ SS &= \sum_{k=1}^n sim_k \end{aligned}$$

$SCORE$  は、入力画像のスコア、 $score_k$  は各類似画像のスコア、 $sim_k$  は各類似画像の類似度、 $SS$  は各類似画像の類似度の和、 $n$  は入力画像に対する類似画像数である。式 (13) では、より高い類似度を持つ類似画像のスコアが入力画像のスコアにより反映される。この加重平均により求めた値を最終的な入力画像のスコアとする。以下、類似画像のスコアを  $score$ 、入力画像のスコアを  $SCORE$  とする。

## 4. 評価実験

提案手法によるスコアリングが人間の主観評価と相関があるかどうかを評価する。今回複数の  $score$  候補式を考案しているため、その中でどの式が一番最適であるかをまず決める必要がある。そこで、前章で挙げた複数の  $score$  候補式の中で、どの式から計算した  $SCORE$  が主観評価と最も相関があるかを調べる。次に、ベストな  $score$  を用いて計算した  $SCORE$  が実際に人間の主観評価と相関があるかどうかを再度同じ方法で評価する。さらに、提案手法の妥当性を調べるため、写真家の写真の撮り方ガイドラインに提案手法が沿っているかどうかを調べる。以降、 $score_{1-10}$  から計算した  $SCORE$  をそれぞれ  $SCORE_{1-10}$  とする。

### 4.1 SCORE 式選定

最も適切な  $score$  は、その  $score$  から計算した  $SCORE$  が人間の主観評価と最も相関がある必要がある。そこで、適切な  $score$  を調査するために、ある画像群内における各  $SCORE$  の順位と、人間の主観評価の順位の相関係数を計算した。今回、Flickr から 10 テーマ × 10 枚の画像を収集し、同じテーマに属する画像群を一つの画像群とした。テーマとしては、「海」「山」「建物」「車」「動物」「バイク」「木」「食べ物」「船」「部屋」の 10 テーマを用意した。ここでは、人間が写っていない画像を収集した。これは、写っている人間の顔や表情に対する好みと、動物の顔に対する好みより、実験参加者によって大きく異なるためである。参加者には、各画像を 10 段階評価 (1: 非常に悪い, 10: 非常に良い) してもらい、テーマごとに参

加者の平均評定と提案手法の各 *SCORE* との相関をケンドール順位相関係数により算出した。最終的な各 *SCORE* の相関係数は、全テーマの相関係数の平均として算出した。本実験における参加者は写真の撮り方について特別な知識を持っていない学生 8 名 (男性 6 名, 女性 2 名) であり, 平均年齢は 22.4 歳である。

表 1 に, 実験で得られた各 *SCORE* に対するケンドール順位相関係数を示す。ただし, *SCORE*<sub>6-8</sub> については,  $\beta = 20, 30, \dots, 100$  のうちの最も相関のあったもののみ記載している。結果として, 式 (14) の加重平均 *SCORE*<sub>2</sub> が人間の主観評価と最も相関があるとわかる。ただし, 式 (14) の第一項の範囲が  $10^{-3} \sim 10^{-2}$  であり, 第二項の範囲が  $0 \sim 10^2$  であったため,  $\gamma = 1/10000$  としている。

$$\text{score}_2 = \frac{\text{favorites}^2}{\text{views}} + \gamma \cdot g(\text{views}, \text{days}) \quad (14)$$

#### 4.2 ベスト *SCORE* の評価

前述の実験より, 式 (14) の加重平均 *SCORE*<sub>2</sub> が, 画像スコアリングに最も適切な式であるという結果が得られた。次に, *SCORE*<sub>2</sub> が本当に人間の主観評価と相関があり, スコアリング式として妥当性があるかどうかを確かめるために再度評価した。前実験と同様に, 10 テーマ  $\times$  10 枚の画像を収集した。収集した画像は前実験で使用した画像と異なる画像であり, 用意したテーマは, 前実験と同様の「海」「山」「建物」「車」「バイク」「木」「食べ物」に加え, 「動物」「船」「部屋」の代わりに「犬」「川」「ボトル」を使用した。前実験と同様の理由で, 使用した画像にはすべて人間が含まれていない。また, 相関係数についても前実験と同様の手順で算出した。本実験における参加者は写真の撮り方について特別な知識を持っていない学生 12 名 (男性 11 名, 女性 1 名) であり, 平均年齢は 22.8 歳である。ただし, 参加者中 5 名は, 前実験にも参加している。

図 3 に, 評価実験で得られた, 各テーマにおける *SCORE*<sub>2</sub> と平均評定のケンドール順位相関係数を示す。ケンドール順位相関係数の全テーマの平均は 0.54 と, *SCORE* 式選定の結果と同等の相関が見られた。関連研究において 0.48 が十分な相関の値としていること<sup>10)</sup>, また, 主観評価との相関であることを考慮すると, *SCORE*<sub>2</sub> と主観評価は高い相関があると言える。ゆえに *SCORE*<sub>2</sub> が画像の魅力を表していると言える。

#### 4.3 写真の撮り方のガイドラインとの比較

提案手法の妥当性を示す第一ステップとして, 提案

手法が写真家の写真の撮り方ガイドラインに沿っているかどうかを検証した。今回は, 提案手法が構図の良し悪しを正しくスコアリングできているかに観点を絞って比較を行った。画像共有サイトを利用するユーザは, 画像にメタ情報を付加する際に, 構図が良いかどうかを考えながら付加しているとは言い切れないが, 今回は試験的に提案手法が構図の良し悪しを評価できているかを評価した。まず, 2 枚の似た画像が良い構図, 悪い構図として紹介されているウェブサイト から, 2 枚の画像ペアを 5 ペア収集した。収集した各画像について *SCORE* を計算し, 各ペアについて, *SCORE* とガイドラインの大小関係が同じかどうかを比較した。その結果, 5 ペアのうち 4 ペアが, ガイドラインの大小関係と一致し, 提案手法の有効性を示すことができた。表 3 に画像とスコアのセットの一部を示す。良い構図と悪い構図のスコアの差の大きさがどれだけの意味を持つか, またスコア値そのものの大きさの意味は, 提案手法のスコアを正規化し, より多くの画像ペアで検証してから議論する必要がある, 今後の検討課題である。

## 5. 考 察

### 5.1 評価実験結果について

まず, *SCORE* 式選定の結果について議論する。*SCORE*<sub>3</sub>, 10 のように掲載日数で割っている *SCORE* と主観評価にはあまり相関が見られなかった。これは, 掲載日数に, 画像が見られていない期間が考慮されておらず, たとえ主観的に見て良い画像であっても, 見られていない期間が長いと, *SCORE* が小さくなってしまいうためであると考えられる。また, *SCORE*<sub>6-8</sub> を比較したときに, *SCORE*<sub>8</sub> の方が比較的相関が見られた。これはお気に入り数がコメント数より人間の主観評価と相関があることを意味する。お気に入り登録という行為はすべてポジティブな行為であるが, コメントはすべてがポジティブなものではなくネガティブなものもあることが反映されていることがわかる。*SCORE*<sub>6-8</sub> の方が *SCORE*<sub>3, 4</sub> より相関が見られるのは, *SCORE*<sub>6-8</sub> はネガティブなコメントを考慮していないが, お気に入り数を *SCORE* に組み込むことでその分を補っているためであると考えられる。これらのことから, コメントの内容を考慮したスコアリングも今後考える必要がある。

次に *SCORE*<sub>2</sub> の評価結果についての考察を述べる。*SCORE*<sub>2</sub> は主観評価と相関があると言える結果

ウェブ上の類似画像に付加されたメタ情報に基づく画像スコアリング

表 1: 各 SCORE の主観評価とのケンドール順位相関係数 ( $\alpha = 100$ , ( ) 内の値は  $\beta$  の値)

Table 1: Kendall's rank correlation coefficient of each SCORE ( $\alpha = 100$ ,  $\beta$  is the value described in ( ))

SCORE1	SCORE2	SCORE3	SCORE4	SCORE5
0.4170	0.5555	0.2860	0.3006	0.3670
SCORE6(30)	SCORE7(30)	SCORE8(20)	SCORE9	SCORE10
0.3995	0.5113	0.5157	0.3805	0.3083

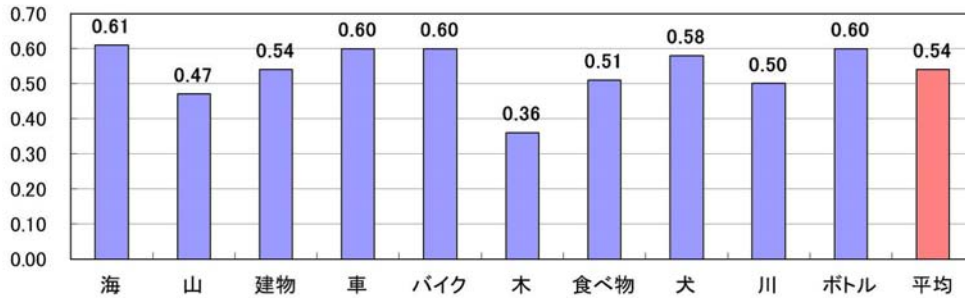


図 3 各テーマにおける SCORE2 と平均評定のケンドール順位相関係数  
Fig. 3 Kendall's rank correlation coefficient for each category

表 3: 提案手法とガイドラインとの比較 (上が良い構図, 下が悪い構図)

Table 3: Comparison between the proposed scoring and photography guideline

(The upper images are good compositions and lower ones are bad compositions according to guideline)

良い構図				
SCORE	0.0538	0.0446	0.0177	0.0062
悪い構図				
SCORE	0.0103	0.0023	0.0076	0.0179

となった。しかし、テーマ「木」については相関があまり見られなかった。これは、入力画像に対する類似画像数が極端に少ない場合があり、スコアリングに必要な情報が十分得られなかったと考えられる。ゆえに入力画像に対する類似画像数が少ない場合のスコアリングを改善する必要がある。一つの解決策としては、類似画像検索の際の類似度の閾値を入力画像によって変更し、より多くの類似画像を取得する方法である。しかし、閾値を変えることで類似画像数は増えるが、

類似度が極端に低い類似画像をスコアリングに使用するのは不適切である。他の解決方法は、独自の類似画像検索を構築することである。今回は類似画像検索を既存の類似画像検索サイト Gazopa に依存しているため、類似度が 40% 以上の類似画像が少ない場合があったが、独自の類似画像検索を構築することで、類似度 40% 以上の類似画像を増やすことができる。

### 5.2 今後の課題

実験結果より、提案手法は魅力に基づく画像スコア

リングとして有効であることが証明されたが、解決すべき問題も存在する。最も重要な問題は類似画像検索が、人間から見て、類似画像と見なせる画像を検索しているかということである。提案手法のスコアリングと相性の良い類似画像検索を利用することで初めて適切なスコアリングを実現できる。今回は、類似画像検索として Gazopa を利用したが、人間が類似していると認識できる類似画像を検索する類似画像検索を実装する必要がある。類似画像検索の改善策としては、タグを利用する方法がある。これは入力画像にタグ付けされているという前提のもと、類似画像検索結果の中からさらに同じタグを持つ画像を検索する手法である。この改善により、状況に応じた画像スコアリングも可能になる。例えば、結婚式という状況の場合は、結婚式特有の魅力的な写真の撮り方が存在すると考えられ、タグを用いて類似画像検索をフィルタリングすることで、結婚式に合ったスコアリングが可能になる。しかし、このフィルタリングを行うためには、画像に写る対象物によって、その対象物特有の魅力的な構図が存在するという仮説を証明する必要がある。

もう一つの問題は、画像共有サイト上の画像に付加されているメタ情報が必ずしも信頼できるとは限らないことである。例えば、主観的に見て良い画像であるにも関わらず、その画像の存在がユーザに気づかれていないために、アクセス数やお気に入り数が 0 に近い場合は、このメタ情報は信頼性がない。もし類似画像としてこのような画像が多く検索されてしまうと、提案手法によるスコアは小さくなってしまふ。ゆえに、信頼できるメタ情報が付加された画像を類似画像として使用する必要がある。Flickr 上の画像は必ずしもすべてが信頼できるメタ情報を持っているわけではないので、独自のデータベースを構築する必要がある。

### 5.3 手法の応用について

提案手法の応用としては、様々なものが考えられる。例えば、提案手法をカメラに搭載して、カメラに写っている構図をリアルタイムでスコアリングすることで、ユーザは、スコアを確認しながら簡単に良い写真を撮影できる。リアルタイムに類似画像検索を行うのは現実的ではないので、まず提案手法により、スコアの高い画像を収集し、収集した画像の特徴量を解析する。解析結果から魅力的である画像であるための特徴量の傾向を決定し、入力画像がその傾向にどれだけ近いかでスコアリングすることが求められる。さらにタグを利用した類似画像フィルタリングにより、状況に応じたスコアリングも可能である。

## 6. 結 論

本論文では、人間の主観評価を考慮できる新しい画像スコアリングを提案した。提案手法では、入力画像に対する画像共有サイト上の類似画像が使用されている。この類似画像はアクセス数、お気に入り数といったメタ情報を持ち、このメタ情報は人間の主観評価と相関があると考えられる。入力画像のスコアは、メタ情報から計算した類似画像のスコアの、類似度を重みとした加重平均として計算される。評価実験では、提案手法と主観評価との相関が実証された。さらに、提案手法が写真家の写真の撮り方ガイドラインに沿うことも証明した。類似画像検索の改善、スコア式の改善、信頼できるメタ情報を持つ新しいデータベースの構築などが今後の課題である。

## 参 考 文 献

- 1) Wee, C. and Paramesran, R.: Measure of image sharpness using eigenvalues, *Information Sciences: an International Journal*, Vol.177, No.12, pp.2533–2552 (2007).
- 2) Obrador, P.: Document layout and color driven image retrieval, *ACM SIGIR Conf. on Research and Development in Information*, pp. 889–890 (2007).
- 3) Zhang, L., Chen, L., Jing, F., Deng, K. and Ma, W.: EnjoyPhoto: A vertical image search engine for enjoying high-quality photos, *Proc. ACM Int. Conf. on Multimedia*, pp.367–376 (2006).
- 4) Zhu, X., Zhao, J., Yuan, J. and Xu, H.: A fuzzy quantization approach to image retrieval based on color and texture, *Proc. Int. Conf. on Ubiquitous Information Management and Communication*, pp.141–149 (2009).
- 5) Belongie, S., Malik, J. and Puzicha, J.: Shape matching and object recognition using shape contexts, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.24, No.4, pp.509–522 (2002).
- 6) Shi, J. and Malik, J.: Normalized cuts and image segmentation, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.22, No. 8, pp. 888–905 (2000).
- 7) Ko, B. and Byun, H.: Integrated region-based image retrieval using region's spatial relationships, *Proc. IEEE Int. Conf. on Pattern Recognition* (2002).
- 8) Chi, M., Zhang, P., Zhao, Y., Feng, R. and Xue, X.: Web image retrieval reRanking with multi-view clustering, *Proc. Int. Conf. on World Wide Web*, pp.1189–1190 (2009).
- 9) Savakis, A., Eitz, S. and Loui, A.: Evaluation of image appeal in consumer photography, *Proc. SPIE Human Vision and Electronic Imaging* (2000).
- 10) Pedro, J. and Siersdorfer, S.: Ranking and classifying attractiveness of photos in folksonomies, *Proc. Int. Conf. on World Wide Web*, pp.771–780 (2009).