Selective Search を用いた構図改善を行う画像トリミングアプリケーションの提案

杉岡広隆[†1]　　水口充[†1]

現在世の中にはスマートフォンやデジタルカメラが普及している. SNS に写真を投稿し,ユーザ間で感動や情報を共有する機会が増えてきている. また,それに伴い自動で写真の魅力を向上させるレタッチソフトが普及してきているが,写真の自動トリミングによって魅力を向上させるアプリケーションは出ていない. 写真の魅力に構図は大きな影響を与える.本論文では Selective Search を利用し,先行研究より適切なトリミングと処理速度の向上を目指した.

1. はじめに

世の中にスマートフォンやデジタルカメラが普及し,今では誰もが写真を撮る事が可能となってきている. それと同時に SNS に写真を投稿し,不特定多数のユーザと情報や感動を共有すると言う事も馴染みのある行為となってきている. 写真は品質が良いだけでは良い写真とは言えない. 表現力を向上させるためにはコントラストやトーンカーブ,ハイライトやシャドウと言った画像のパラメータを調節する事が有効となる.そのため SNS に写真を投稿する事が普及してくると同時に,撮った写真を加工する Adobe Photoshop[1]などのレタッチソフトも普及してきた. Instagram[2]などのスマートフォン SNS アプリ自体にも特殊効果を付与するレタッチ機能が備わっている. このことからも写真をより魅力的に見せるアプリケーションは需要があると言える.

写真の品質や特殊効果以外にも表現力を向上させ,より魅力的に見せる要素の一つに構図がある. 構図は写す主題(被写体)を明確にし,その被写体の配置と空間のバランスを考慮しなければならない. その為には構図の知識や被写体に適した構図を決定する判断力と言った専門知識が必要となる. 被写体が決まっていても上手く写真が撮れない場合には構図に問題がある事が多いと言われている[3]-[12].
画像に自動的に適した特殊効果を付与するアプリケーションは多く存在するが,構図は昔と変わらず今も自動化はされていない[13].撮影者(画像製作者)自身が行わなければならない. そのため,専門知識の無いユーザでも構図の考えられた撮影が行えるよう,日の丸構図や三分割構図などの 5 つの構図のいずれかに適した撮影領域へ誘導を行う撮影支援の研究[14]や,カメラの画面上に構図テンプレートを映し出す撮影支援アプリケーション[15]などが開発されている. しかし複数の構図テンプレートが用意されているとはいえ,ユーザがその中から適切なものを選択する必要がある事や,用意されていない構図に対応できない点,複数の構図を適応した提案構図について考慮されていない. と言った課題点がある.

他に,撮影後の写真に三分割構図が適応されているか否かを判定する手法を提案した研究[16]がある. 画像を良く見せるため,最適化問題を粒子群アルゴリズムで解いた,切り取りと折り返しによる手法の提案[17]もされた. しかしこれは計算量が多く,切り取りでは元の情報が失われ折り返しでは不自然に線対称模様が生じてしまうことが問題であった. そこで画像の平均化やガウシアンフィルタなどを利用し,それらの差から得た勾配強度を重要度とし,そこを元に画像をリサイジングする手法[18]が提案されたが,被写体を部分的にしか検出しないといった問題点があった. 視線追跡を用いて,視線から重要度を判定し画像から被写体を切り出す研究[19]があるが,視線追跡を行う手間がかかる.

以上の様に,画像をより魅力的に見せる事には需要があり,その要素の中に品質・特殊効果・構図がある. 品質はカメラの性能に左右され,特殊効果を付与するアプリケーションは多々存在する. しかし,撮影時の構図を考慮する支援アプリケーションは存在しても,撮影後の画像をより魅力的に見せる,トリミングを行うアプリケーションは発展途上であると言える. そこで本研究では,R-CNN の物体候補(Selective Search)アルゴリズムと,K-means 法を利用して検知した物体の重要度を決定し,複数の構図から適した構図を選択,トリミングを行う手法を提案する.

2. 被写体の検出

構図を決定する要素として被写体の画像上での位置,大きさ,数を利用する. そのため構図の改善を行う前に,まず画像の主題となる被写体を選定する必要がある. 以下では画像に対して物体検出と重要度の設定による被写体の選定を行う.

†1 京都産業大学大学院
Kyoto Sangyo University Graduate School

ⓒ2021 Information Processing Society of Japan

(1) Selective Search を利用した物体検出

ニューラルネットワークに畳み込みを導入した画像の畳み込みを表現する CNN(Convolutional Neural Network)は非常に処理が遅いなどの問題がある．その問題を解決したのが物体候補(Selective Search)アルゴリズムを使用した R-CNN(Region-based Convolutional Neural Network)である[20].

本研究では,R-CNN で利用される Selective Search を用いて物体検出を行う．以下の図 1 は Selective Search の処理過程を表したものであり,画像上の①～④はそれぞれ対応している．

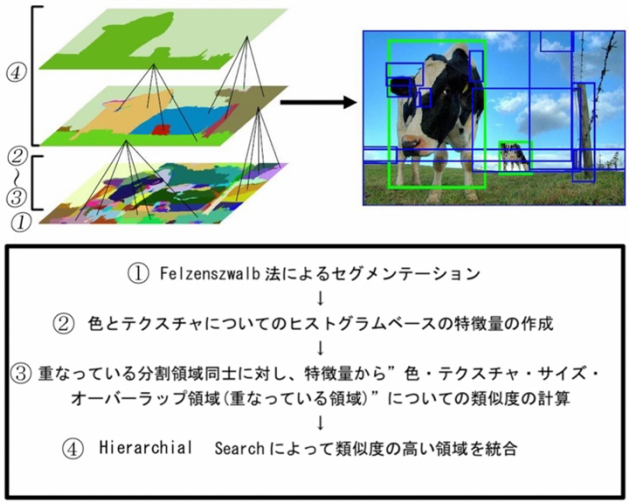

図 1 Selective Search の処理の概要

Selective Search は,図 1 の様に色やテクスチャの類似度の高い近接領域を段階的に結合し,最終に残った領域を画像に写った物体であるとする手法である．

Selective Search を用いて物体検出を行なった結果例が以下の図 2 である．なお縦もしくは横に極端に細長い矩形や小さすぎる矩形はノイズとなるため,矩形の縦横の長さの積が 2000pix 以上になり,縦横の比率が 1.4(白銀比)以上にならないよう閾値を設けている．閾値の 2000pix は,1700*1133pix の元画像に対して 100 から順に 100 ずつ上げていった際に最も精度が良いと感じた値である．

図 2 を確認すると被写体である鳥を囲う矩形が存在する事が確認できる．次に,この多く表示されている矩形のうちいずれが写真の主題(被写体)を表しているのか,被写体はいくつあるのか,被写体の画像空間上での位置は何処であるのか,大きさはどれ程であるのか．これらを決定する必要がある．

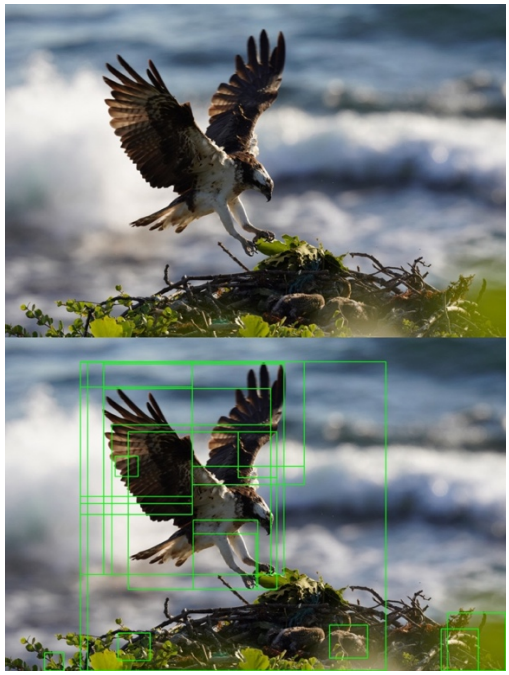

図 2 元画像 (https://www.sony.jp/ichigan/a-universe/news/351/) と Selective Search による物体検出結果

(2) K-means 法による被写体の選定

矩形が多く表示されている場所は重要度が高いという仮説を考え,より多く矩形が固まっている地点を発見する．

本研究では内部処理を早くしたいと考えているため,実装が容易で階層的クラスタリングに比べ計算量の少ない,クラスタの平均を用いてクラスタを調整する非階層的なクラスタリングの手法の一つである K-means 法[21]-[23]による被写体の選定を行う．

まず検出した矩形の中心点座標(特徴点)の座標を平均が 0・標準偏差が 1 となる様に標準化する．クラスタリングの結果,クラスに含まれる特徴点が最も多いクラスの矩形集合をメインの被写体とし,メインの被写体であると判定されたクラスの特徴点数の 80%以上の特徴点数(同程度の重要度)を持つ他のクラスも被写体クラスであるとする．また K-means 法はクラスタの数を定数として予め定義する必要がある．ヒトが同時にいくつの物に注意をする事が出来るのか,という分割的注意について,実験心理学では複数物体追跡課題を用いて検証が行われている．その結果,ヒトは約 4 個の物に対して同時に注意を向ける事が可能であると言われている[24]-[25]．この事から,ヒトが画像を見た際に注意を向ける場所は多くて 4 箇所であると考えられる．よってクラスタリングにおけるクラスの数の最大を 4 とし,特徴点の数とクラス数の対応を以下の表 1 のように定めた．K-means 法を用いてクラスタリングを実行した結果が以下の図 3 である．

表 1 特徴点とクラス数の関係

クラス数(個)	1	2	3	4
特徴点(個)	10 以下	20 以下	30 以下	41 以上

ⓒ2021 Information Processing Society of Japan

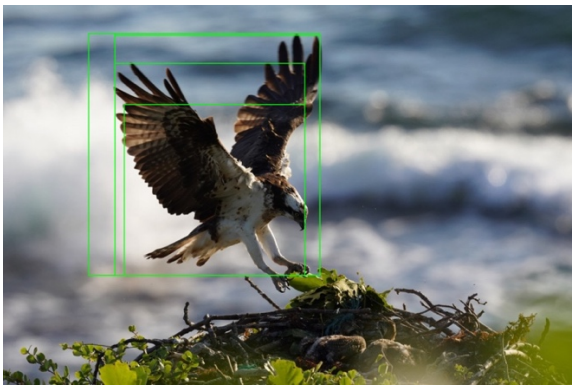

図 3 K-means 法を用いたクラスタリング結果

図 3 では被写体である鳥を囲む矩形のみが表示されている事が確認できる．ここで,検出結果に対して物体検出の評価指標である IoU(Intersection over Union)を求める計算を行う．IoU は以下の式で求める事が可能である．

$$IoU = \frac{\text{Intersection(領域の共通部分)}}{\text{Union(領域の和集合)}} \quad (1)$$

元画像の鳥は座標(330 ,84)-(952 ,856)に存在し,検出結果は(255 ,82)-(1015 ,805)であった。結果,IoU ≒ 0.77 であった．IoU は 2 つの同じ正方形を斜めに 1/9 ずらすだけでも約 0.65 まで落ちる指標であるが,0.77 は十分に高く,適切に物体検出が行えていると言える．

3. 構図

3.1 構図の種類

前章で検出した被写体の数と大きさ,位置といった構成要素の関係から適切な構図を決定し,構図を反映させるようトリミングを行う．写真撮影で利用される構図は多種多様であるが,数多くある構図の中でも多くの本や論文,ウェブページの半数で紹介されている構図の基本となると考えられているもの[3]-[14]の中でも有名な日の丸構図・三分割構図・対比構図・対角線構図の 4 つの構図(図 4)を入力画像に適応させる．なお図 4 の赤い丸は被写体を配置する位置を示している．

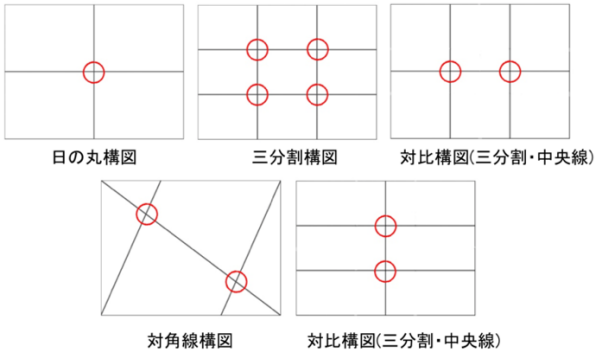

図 4 実装する構図パターン

3.2 日の丸構図

検出した被写体の中点が画面の中心に来るようにトリミングを行う．しかし大きくトリミングして多くの画素を切り捨ててしまうと画質(解像度)が落ちてしまい,写真をよりよく見せる要素の一つである「品質」が損なわれてしまうため,日の丸構図を実現できる最小のトリミングを行う．また,日の丸構図は検出された被写体の領域が元画像の 1/2 以上である場合に実行する．日の丸構図が適応された例が以下の図 5 である．

図 5 日の丸構図の適応例

3.3 日の丸構図

三分割構図とは三分割線の交点のいずれかに被写体が配置される構図である．三分割構図は被写体が一つであり,その被写体領域の面積が元画像の 1/2 未満の場合に適応させる．三分割構図が適応された例が以下の図 6 である．また,領域の大きさが 1/2 未満の被写体が三つ以上存在した場合は,出力画像におけるフォーカルポイントが多く,更に画像を占める被写体の面積が大きくなり三分割構図の特徴である背景の広がりを感じさせる効果が損なわれると考える．これらの事から被写体が二つ以上存在した場合は日の丸構図を適応する．

図 6 三分割構図の適応例

ⓒ2021 Information Processing Society of Japan

3.4 対角線構図

写真内に被写体が二つ存在した場合,二つの被写体領域を囲う領域を一つの被写体領域として定義する．その新たな領域を用いて日の丸構図を適応させる事で対角線構図となると考える．しかし,この方法は二つの被写体が対角線上に並んでいることが前提である．対角線構図の基準位置との差の許容範囲は±8.5°であるという調査結果がある[14]．そのため画像の対角線と,二つの被写体を結ぶ線の角度に±8.5°より大きな差がある場合は二つの被写体は対角ではなく横もしくは縦に並んでいると見なし,対比構図を適応する．対角線構図が適応された例が以下の図 7 である．

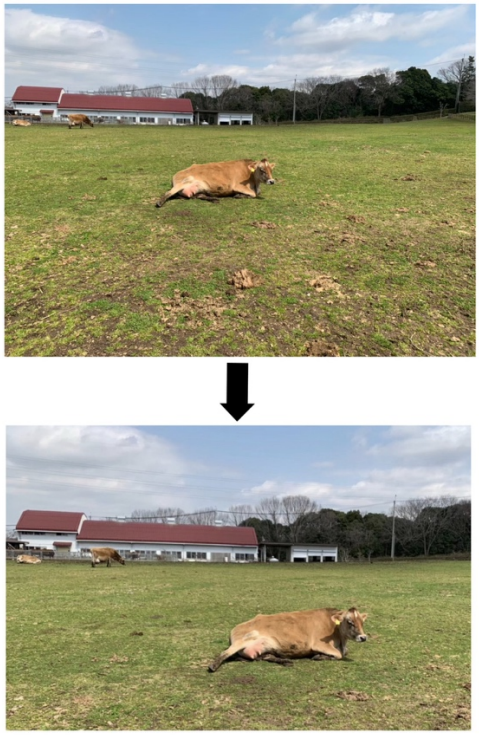

図 7 対角線構図の適応例

3.5 対比線構図

被写体領域の幅と高さのどちらかが元画像の 1/2 以上,もう片方が 1/2 未満であり,被写体領域面積が元画像の 1/2 未満の場合に対比構図を適応させる．図 4 の対比構図のテンプレートは二分割構図と三分割構図のハイブリッド構図である．そのため,二分割構図と三分割構図に振り分ける条件を半分ずつ満たしている被写体にこの構図を適応させる．また,前章で被写体が縦方向か横方向に並行に並んでいると判定された,被写体を二つ有する画像にも対比構図を適応する．対比構図が適応された例が以下の図 8 である．

図 8 対比構図の適応例

4. 実験

4.1 実験目的

出力結果が入力画像よりユーザにとって好ましいものとなっているのかといった観点から有効性を調査する．また提案システムを利用する方が専門知識を有していない写真のアマチュアがトリミングする場合より好ましい出力をするのか否かという観点から有用性を調査し,提案システムの評価を行う．

4.2 実験手順

写真撮影の知識の乏しい 18 歳～26 歳のアマチュア 10 名が撮影した写真を,それぞれから 2 枚ずつ計 20 枚提供してもらった．それぞれに提供してもらった 2 枚の写真のうち 1 枚を撮影者本人の手で見栄えが「より良く」なるようにトリミングするように指示した．その際,トリミングが必要ないと感じた場合はトリミングを行う必要はないと伝えたが,10 名ともトリミングを実行した．次に,残った 10 枚の写真に対し本研究の提案システムを実行した．

提供してもらった元画像と提案システムを通し「トリミングが行われた画像のペア」と「元画像と提供者自らトリミングを行なった画像のペア」の二種類の 20 ペアを被験者に提示する．

二枚の写真のペアを提示された被験者には,直感でどち

ⓒ2021 Information Processing Society of Japan

らが「いいね」と感じたのかを回答してもらった．実験の前に被験者には,どちらを選択しようとも正解はないため,直感で選択するよう指示をしている．この実験の対象者は10 代前半から 30 代前半の男女 302 名である．画像の提示と選択の実験は Google forms によって行われ,被験者各々の好きなデバイス上で実行する．デバイスを指定していないのは,本研究の提案システムは SNS に画像を投稿するための画像加工支援を目的としており,SNS は PC やスマートフォンなど様々なデバイスで利用されるため,デバイスや画面のサイズ,色合いを指定するべきでないと考えたためである．

4.3 結果

被験者には 2 枚の画像のうち「いいね」と直感で感じた方を回答してもらったが,トリミング後の画像を”1”・元画像を”0”として収集したデータを分析する．提案システムがトリミングした画像の選択率と,撮影者本人がトリミングを行なった画像の選択率はそれぞれ以下の表 2 のようになった．

表 2 結果

標本	提案システム	撮影者
平均	0.660	0.453
分散	0.012	0.018
観測数	10	10
自由度	9	9
t	4.396	
P(T<=t) 片側	0.001	
t 境界値 片側	1.833	
P(T<=t) 両側	0.002	
t 境界値 両側	2.262	

一般的に $p<0.05$ の状態で「有意差有り」と判断される[26]が,表 8 の「P(T<=t) 両側」の値を確認すると,0.002 と 0.05 以下になっている事から,2 群のデータが等しいという仮説の域に落ちないために帰無仮説は棄却され有意差が認められる．p の値が 1%未満となっている事から結果が偶然である可能性は低く,両標本に差があるという対立仮説が正しいと考えられる．よって,提案システムを利用したトリミング画像の方が優位に選択されていると言える．

4.4 提案システムの有効性

表 2 の提案システムの選択率の平均値を確認すると 66%とおよそ 7 割の人が提案システムを通した画像の方が良いと回答している．この事から提案システムを利用する事で入力画像がよりユーザにとって好ましいものとなり出力され,効果的に働いていると考えられる事から,有効性があると考える．提案システムがトリミングを行なったが,その画像の選択率が 50%を超えていな画像が 1 枚あった．この画像の評価が低い理由として,トリミング範囲の過剰縮小による画質の低下によるものと,システムの適応できる構図の他により良い構図があり対応できていなかった事が原因であると考える．

4.5 提案システムの有用性

撮影者がトリミングを行なった画像の選択率の平均は表 2 より 45.3%であり,50%を下回っている事が分かる．専門知識を有していない写真のアマチュアがトリミングを行うよりも提案システムを用いた方がより良いトリミングを行う事が可能であると言え,提案システムに有用性があると言える．

5. 今後の展望

本研究で作成した提案システムでは,日の丸構図・三分割構図・対角線構図・対比構図の 4 つの構図のいずれかを入力画像に適応させていたが,構図の種類が少ないために入力画像に不適切な構図を適応させてしまう場合がある．適応可能な構図の種類を増やすため,画像に含まれる線分から地平線や消失点・画像の傾きを検出することで,角度修正や地平線・放射線を用いた構図が実現できるのではないかと考える．また,S 字や三角構図と言った画像要素内に含まれている図形を用いた構図はテンプレートマッチングによって検出可能なのではないかと考える．

提案システムでは物体検出を行なった後に k-means 法によってクラスタリングを行なっていた．k-means 法ではクラスタの数が定数であるため,特徴点の数に応じてクラスタ数を変化させていた．被写体の選定はおおよそ適切に行われていたが,表 1 の特徴点数とクラスタ数の関係の妥当性は証明されていないため,妥当性を検証する必要がある．

提案システムによる出力画像は 4 つの構図のうち 1 つを適応させた 1 枚のみであったが,複数の構図を適応する事が可能である場合も存在するためその場合は,ユーザに複数の出力結果を提示し,好ましと感じるものを選択する操作を与えることで,あたかも自分がトリミングを行なったかの様な満足感を得る事ができるのではないかと考えた．また複数の構図を提示されてもユーザによって感性が異なる為,提示された構図の中で迷い,選択しきれないと言った場合もあると考えられる．構図適応の処理内に「優先度」と言ったパラメータを設け,どの構図でどの処理手順によって適応させたのかと言った情報を元に優先度を計算する．ユーザが提示された複数の構図画像から選択しきれない場合,優先度が高く良いと考えられる構図(おすすめ)順に並び替えると言った様な機能があると良いと考える．

提案システムでは入力画像の情報を小さくする事などはせず,そのまま処理を行なっていた．システムに入力された画像データの画質を落とし,軽くなったデータに対して計算処理を行い,その後に元画像にトリミング処理を実行する事によって処理の高速化が望めるのではないかと考える．

本研究では SNS に投稿するための写真の加工処理支援を目的としてシステムを作成した．写真では構図が重要であるが,イラストでも同じ事が言える．SNS に自ら描いたイラ

ⓒ2021 Information Processing Society of Japan

ストを投稿するユーザも多く存在するため,このシステムをイラストの構図デザイン支援ツールとして利用できないかと考えた.

6. まとめ

先行研究には被写体検出の最適化問題を粒子群アルゴリズムで解いた手法[17]や画像の勾配強度を重要度とし,被写体を特定する手法[18],目線から画像の部分的重要度を図る手法[19]があるが,粒子群アルゴリズムは計算に時間がかかってしまい,勾配強度を用いた被写体検出方法は被写体の一部しか検出できず,目線を利用するのは手間がかかると言った事が問題であった. しかし,被写体の検出に R-CNN の selective search と k-means 法を用いる事で 4000*3000pix 程度の画質の入力でも 10 秒程度で手間をかけずに出力を行い,なおかつ被写体全体を覆う様に被写体領域を指定する事が可能となった.

構図の適応方法として,とにかく構図を適応させるわけではなく画像の品質(画質)を保つため,極力画素を切り落とさず元画像の印象を保ったまま,目的の構図となる最大の画像サイズになる様にトリミングしている.また,処理は AI や深層学習・機械学習と言った複雑な処理かつブラックボックス化されているものや,複雑な計算は利用せず大半は四則演算のみで構図が適応されている.

適応できる構図が 4 つと数が少ないためにテンプレートマッチングなどを利用し構図の種類を増やす他,画像の傾き補正・回転などを利用して柔軟に多くの構図に対応していきたい. またユーザに複数の出力結果を提示し,複数ある中から最終出力結果を選択してもらう事で満足度が上がるのかを検証していきたい.

参考文献

1) Adobe Photoshop, https://www.adobe.com/jp/products/photoshop.html（最終閲覧日 2021 年 7 月 25 日）
2) Instagram, https://www.instagram.com（最終閲覧日 2021 年 7 月 25 日）
3) 桜井始:すぐに役立つ被写体別画面構成, 玄光社, (2008).
4) 山口高志: 風景写真被写体別構図の決定, 学習研究所, (2008).
5) 山口高志:写真撮影とらの巻フレーミング実例辞典, 学習研究所, (2008).
6) デジタルフォト編集部: 今さら人に聞けないデジタル一眼レフの疑問構図編, ソフトバンククリエイティブ, (2009).
7) 大和田良: 三分割構図法を完璧にマスターする, デジタルフォト 2009 年 12 月号, pp 41-76, ソフトバンククリエイティブ, (2009).
8) 内田広由紀: 構図エッセンス, 視覚デザイン研究所, (2006).
9) 荻窪圭: デジカメ撮影のネタ帳シーン別ベストショットの撮り方,エムディーエヌコーポレーション, (2009).
10) 西川晃太郎: キレイに見せるデジカメ写真術手軽に楽しむ撮影とフォトレタッチ, ソフトバンククリエイティブ, (2008).
11) 山口高志: 写真撮影フレーミングの決定, 学習研究所, (2009).
12) 岡嶋和幸: デジタル一眼レフ初心者の為のレタッチの基本教えます, 玄光社, (2007).
13) 山田芳文: 写真は「構図」でよくなる, エムディーエヌコーポレーション, (2018).
14) 板宮古宏, 御手洗紘子, 古高淳夫: 構図と顕著性に基づく写真撮影支援手法に関する研究, ITE Technical Report Vol 37, NQ. 12 HI2013 −52, 3DIT2013 −20, (2013).
15) センサカメラ, https://play.google.com/store/apps/details?id=com.wilfredo.bigol.sensorcamera&hl=jp (最終閲覧日 7 月 25 日).
16) Long Mai, Hoang Le, Yuzhen Niu, Feng Liu ; “ Rule of Thirds Detection from Photograph ”, IEEE ISM 2011, pp.91−96, (2011).
17) L.Liu, R.Chen, L.Wolf, D.Cohen-Or: Optimizing photo composition, Comput., Graph. Forum, 29,2, pp.469-478, (2010).
18) 傳暁宇, 井上光平, 浦浜喜一: 画像の分割直交李サイジングによる写真の構図改善, 映像情報メディア学会誌 2010 年 64 巻 12 号 p.1960-1963, (2010).
19) A.Santella, M.Agrawala, D.DeCarlo, D.Salesin, and M.Chen.: Gaze-based interaction for semi-automatic photo cropping in Proc. SIGCHI Conf.Human Factors in Computing Systems, pp 771-780, (2006).
20) Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik, UC Berkeley: Rich feature hierarchies for accurate object detection and semantic segmentation Tech report (v5), (2014).
21) Steinhaus, H: Sur la division des corps matériels en parties (French). Bull. Acad. Polon. Sci. 4 (12): 801–804, (1957).
22) E.W.Forgy: Cluster analysis of multivariate data: efficiency versus interpretability of classifications, Biometrics 21: 768–769, (1965).
23) MacQueen,J.B : Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press. pp. 281–297, (1967).
24) 岩波書店: 科学 2007 年 3 月号 Vol.77 No.3 連載ちびっこチンパンジー 第 63 回「複数のものを同時に見る」, (2007).
25) 早坂太一: 複数物体追跡課題における物体数, 運動方向および速度の影響, 豊田工業高等専門学校研究紀要 第 42 号, (2009).
26) 山田剛史, 村井潤一郎: よくわかる心理統計, ミネルヴァ書房, (2004).

ⓒ2021 Information Processing Society of Japan