

## 視覚的特徴と顔検出に基づく被写体の抽出と分類

Extracting and categorizing photographic subject  
based on visual features and face detection永留 壮馬†  
Soma Nagadome青木 茂樹†  
Shigeki Aoki宮本 貴朗†  
Takao Miyamoto

## 1. まえがき

近年、デジタルカメラの普及によりユーザが大量の写真撮影し、コンピュータ等に保存できるようになった。ユーザによって撮影・保存された写真は膨大な枚数となるため、整理する作業量が膨大となっている。そこで、写真を自動的に整理し提示する技術が注目されている。写真の提示に関する従来研究として、文献 [1] が提案されている。文献 [1] では、写真中の注目領域から抽出したコントラストやぼけ等の撮影技術に関する情報を基に、高品質な注目領域(被写体)を抽出している。この手法では、写真から被写体を抽出し品質を評価することができるため、高品質な被写体をユーザに提示することができる。

本稿では、エッジ・色・ピンぼけの有無等の視覚的特徴を用いて写真から被写体を抽出し、さらに抽出した被写体が人物か物体かを分類する手法を提案する。本手法では、被写体を抽出するだけでなく被写体を分類しているため、写真整理の自動化手法への応用が容易であると考えられる。

## 2. 被写体の抽出

被写体とは、一般にユーザが意図して撮影した人物または物体のことであり、ユーザは以下の三つの性質を満たす様に撮影すると考えられる。そこで、以下の三つの性質を満たしている領域を被写体として抽出する。

- (1) 背景との違いが明確である。
- (2) 人の注意を引きつける色合いを持つ。
- (3) ピンぼけしていない。

ここで、風景画等の写真は被写体と背景の違いが明確でなく、写真全体に人の注意を引きつける色合いを持つため、上記の性質を満たさないと考えられ、別の特徴を定義する必要がある。そこで本研究では、写真全体が被写体の写真を対象外とする。

## 2.1. 被写体候補領域の抽出

まず、写真中に背景との違いが明確な領域が存在するかを確認するために、エッジを抽出する。

入力画像に  $x, y$  方向の  $3 \times 3$  の sobel フィルタを畳み込み、画素  $i$  ごとにエッジ強度  $e_i$  を取得する。そして、画像をラスタスキャンし、エッジ強度の分散  $v$  を算出する。その後、次式により被写体候補度  $C_{e,i}$  を算出する。

$$C_{e,i} = 100(1 - \exp(-\frac{e_i^2}{2v})) \quad (1)$$



(a) ロングショット



(b) アップショット

図 1: 撮影方法ごとの写真の例

次に、人の注意を引きつける色情報を持つ領域が存在するかを確認するために、領域の色情報を抽出する。被写体から得られる色情報は撮影方法によって異なると考えられる。ユーザが被写体をロングショットで撮影すると、図 1(a) に示すように被写体の面積が小さくなるため、被写体がつ色の出現頻度は低くなる。一方、被写体をアップショットで撮影すると、図 1(b) に示すように被写体の面積が大きくなるため、被写体がつ色の出現頻度は高くなる。そこで、色の出現頻度に注目する。

まず、表色系を変換する。表色系には、一般に人の色の知覚に類似していると言われる HSV 表色系 (H:色相, S:彩度, V:明度) を用いる。その後、HS 平面を  $10 \times 8$  の領域に分割し、各画素の H, S 値から色の出現頻度を算出し、被写体候補度を求める。色の出現頻度の低さを表す被写体候補度  $C_{d,i}$  を式 (2)、色の出現頻度の高さを表す被写体候補度  $C_{u,i}$  を式 (3) によって算出する。

$$C_{d,i} = \left(1 - \frac{b(h,s) - \min_{h,s} b(h,s)}{\max_{h,s} b(h,s) - \min_{h,s} b(h,s)}\right) \times 100 \quad (2)$$

$$C_{u,i} = \frac{b(h,s) - \min_{h,s} b(h,s)}{\max_{h,s} b(h,s) - \min_{h,s} b(h,s)} \times 100 \quad (3)$$

ここで、画素  $i$  の色情報  $h, s$  に対応する HS 値の頻度 (ビン) を  $b(h, s)$  で表し、 $\max_{h,s} b(h, s)$ ,  $\min_{h,s} b(h, s)$  はビンの最大・最小値である。

最後に、式 (1)、式 (2) と式 (1)、式 (3) の被写体候補度をそれぞれの和により統合し、算出したそれぞれの被写体候補度がしきい値以上の画素をラベリングして、被写体候補領域として抽出する。このとき、ラベルの面積・幅・高さがしきい値以下ならば、そのラベルを被写体候補領域から除去する。図 2(a)、図 3(a) に示す入力画像から被写体候補領域を抽出した例を図 2(b)、図 3(b) に示す。

†大阪府立大学大学院人間社会システム科学研究科

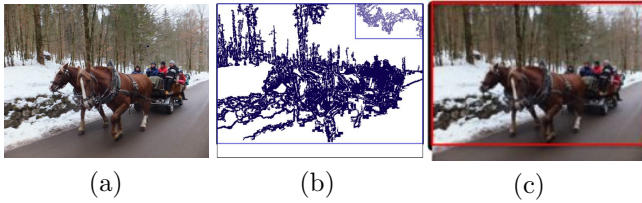


図 2: 式 (1),(2) による被写体の抽出

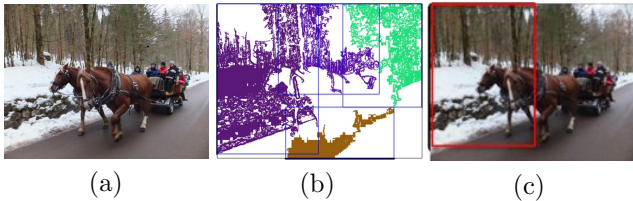


図 3: 式 (1),(3) による被写体の抽出

## 2.2. 合焦点に基づく被写体抽出

ここでは, SVM(Support Vector Machine) によってピンぼけの有無を識別する. まず, ピントの合っている画像に 1, ピントの合っていない画像に 0 を与えて SVM で学習する. その後, 入力画像を  $64 \times 48$  画素にブロック分割し, SVM を用いてピントが合っているか否かをブロック単位で識別する. 次に, 2.1 で抽出した被写体候補領域中の全画素について, ピントが合っている画素数  $N_f$  と合っていない画素数  $N_d$  をそれぞれ数え,  $N_f/(N_f + N_d)$  の値を合焦点として算出する. そして, 全ての被写体候補領域で合焦点を算出し, 合焦点の最も高い領域を被写体として選択する. 図 2(c), 図 3(c) に被写体候補領域から被写体を選択した例を示す.

## 3. 被写体の分類

被写体が人物であるか物体であるかを分類するために, 抽出した被写体に対して文献 [2] の手法を用いて顔を検出する. さらに, 検出された顔の外接矩形領域から肌色の色相 (H 値) を抽出し, 肌色色相の画素数  $N_b$  を数える. その後, 外接矩形領域中の肌色色相の割合を  $N_b/F_r$  によって算出する. ここで,  $F_r$  は外接矩形領域の面積である. そして, 算出した割合がしきい値以上のとき, 被写体を人物と分類する. 一方, 算出された面積比がしきい値未満のとき, 被写体を物体と分類する.

## 4. 実験と考察

本手法の有効性を確認するために, 人物が被写体の写真 150 枚と物体が被写体の写真 150 枚の計 300 枚の写真を用いて被写体を抽出し分類する実験を行った.

まず, 被写体候補領域の抽出実験を行った. 実験では, ユーザが撮影したいと考えた物体や人物を被写体の正解データとして与え, 抽出した被写体候補領域に正解データが含まれている割合を調べた. 実験結果を表 1 に示す. 実験の結果, 72.7% の写真において被写体候補領域の抽出に成功した. 背景が複雑でない写真の場合, 抽出した被写体候補領域に被写体のみが含まれていたが, 背景が複雑な場

表 1: 被写体候補領域の抽出結果

-	被写体のみ	被写体と背景混在	被写体を含む領域
物体	62 (41.4%)	54 (36.0%)	116 (77.3%)
人物	26 (17.3%)	76 (50.7%)	92 (68.0%)
物体+人物	88 (29.3%)	130 (43.3%)	218 (72.7%)

表 2: 被写体の抽出結果

物体	人物	物体+人物
94 (81.0%)	76 (74.5%)	180 (78.9%)

表 3: 被写体の分類結果

		実験	
		物体	人物
定義	物体	78	16
	人物	23	53

合は被写体と背景が混在していた.

次に, 被写体の抽出実験を行った. 実験では, 抽出した被写体候補領域から合焦点に基づいて被写体を選択し正解データと比較した. 結果を表 2 に示す. 78.9% の写真において被写体の選択に成功した. 被写体だけにピントが合い背景にピントが合っていない写真では, 正しく被写体を選択できていたが, パンフォーカスの写真では正しく被写体を選択できていなかった.

最後に, 被写体の分類実験を行った. 分類実験では, 選択した被写体が人物であるか物体であるかを分類し正解データと比較した. 実験結果を表 3 に示す. 再現率 69.7%, 適合率 76.8% の結果が得られた. 正面を向いている人物や肌色色相が少ない物体等は正しく識別できていたが, 下を向いている人物等では分類に失敗していた.

## 5. まとめ

本稿では, 写真の自動整理への応用を目指して, エッジと色の出現頻度とピンぼけの有無等の視覚的特徴に基づいて被写体を抽出し, 顔検出と肌色検出によって被写体を分類する手法を提案した. 実験の結果, 被写体の抽出率は 78.9% であり, 被写体を 72.8% 正しく分類できた. 今後の課題として, 写真の類似色に注目した特徴等を用いて, さらに被写体の抽出率を向上させることが挙げられる.

## 参考文献

- [1] 西山 正志, 岡部 孝弘, 佐藤 洋一, 佐藤 いまり, “複数の注目領域を用いた写真の主観的品質の識別,” 信学論, Vol.J93-D, No.8, pp1364-pp1374, 2010.
- [2] Paul Viola, Michael Jones, “Rapid Object Detection using a Boosted Cascade of Simple Features,” Proc. of CVPR '01, pp511-pp518, 2001.