

HSV 色空間における座標と PSPNet のラベルを利用した風景画像検索

興津祐人 長名優子

東京工科大学 コンピュータサイエンス学部

1 はじめに

これまでに様々な手法を用いた風景画像検索に関する研究が行われてきているが、近年では畳み込みニューラルネットワーク [1] を利用した風景画像検索に関する研究も盛んに行われている。そのような手法の一つとして画像分類を学習した畳み込みニューラルネットワークである VGG16[2] を用いて画像の特徴ベクトルを抽出し、それを用いて類似画像検索を行う手法 [3] が提案されている。この手法では、画像を学習済みの VGG16 に入力し、出力層の 1 つ手前の層の出力を特徴ベクトルとして用いる。画像と特徴ベクトルを対応付けてデータベースに登録しておき、データベース内の画像の特徴ベクトルとキー画像の特徴ベクトルを比較することで検索を行う。この手法では、画像に含まれているオブジェクトが類似していれば特徴ベクトルも類似したものになると考えてこのような方法で検索を実現している。しかしながら、この手法ではどこにどのようなオブジェクトが含まれているかを正確に考慮した検索は実現できていないという問題がある。

一方で、PSPNet[4] という画像分割手法が提案されている。PSPNet は畳み込みニューラルネットワーク [1] を用いた画像分割手法であり、画素単位での画像認識と領域分割を同時に実現することができる。PSPNet による画像認識結果を利用することで、画像のどの位置にどのようなオブジェクトが含まれているかを考慮した検索を行える可能性がある。

そこで本研究では、HSV 色空間における座標と PSPNet のラベルを利用した風景画像検索を提案する。従来手法 [3] のどこにどのようなオブジェクトが含まれているかを考慮できないという問題に対しては、PSPNet のラベルを用いることで解決する。

2 PSPNet

PSPNet (Pyramid Scene Parsing Network)[4] は Encoder に ResNet101[5] の畳み込み層を利用したネッ

トワークであり、Encoder と Decoder の間に空間ピラミッドプーリングモジュールを追加した構造を持つ。画像を入力すると画像に存在するオブジェクトに対して画素単位でラベリングを行い、画像分割と認識を同時に実現することができる。画像分割と認識の結果はのようにラベルごとに色分けされた画像として出力される。

3 HSV 色空間における座標

ここでは、提案手法で利用する HSV 色空間における座標について説明する。HSV は、色相 (Hue: H)、彩度 (Saturation: S)、明度 (Value: V) で色を表現する表色系である。

HSV を座標に割り当てたものが HSV 色空間であり、提案手法では円錐形で表される HSV 色空間を用いる。色相は具体的な色を決める要素であり、0~360度の角度で表される。彩度は色の鮮やかさや濃さを決める要素であり、0~1の範囲の値をとる。明度は色の明るさや暗さを決める要素であり、0~1の範囲の値をとる。画素 i の HSV の値 h_i, s_i, v_i は画素 i の RGB の値 r_i, g_i, b_i ($0 \leq r_i, g_i, b_i \leq 1$) から次のように求められる。

$$h_i = \begin{cases} \frac{60(g_i - b_i)}{A - B} & r_i = A \\ \frac{60(b_i - r_i)}{A - B} + 120 & g_i = A \\ \frac{60(r_i - g_i)}{A - B} + 240 & b_i = A \end{cases} \quad (1)$$

$$s_i = \frac{A - B}{A} \quad (2)$$

$$v_i = A \quad (3)$$

$$A = \max(r_i, g_i, b_i) \quad (4)$$

$$B = \min(r_i, g_i, b_i) \quad (5)$$

HSV 色空間は、図 1 に示すような色空間である。HSV は色相、彩度、明度を用いて色を表現する方法であり、提案手法では、これを円錐型で表される HSV 色空間に写像したときに座標を特徴量として用いる。

Landscape Image Search using Coordinates in HSV Color Space and PSPNet Label
Yuto Okitsu and Yuko Osana (Tokyo University of Technology, osana@stf.teu.ac.jp)

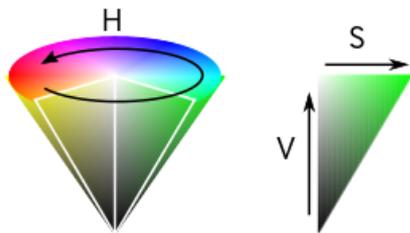


図 1: HSV 色空間の座標

HSV 色空間における座標は色相、彩度、明度の値から

$$\begin{aligned} x_i &= s_i v_i \cos(h_i) \\ y_i &= s_i v_i \sin(h_i) \\ z_i &= v_i \end{aligned} \quad (6)$$

で与えられる。ここで、 h_i は画素 i の色相、 s_i は画素 i の彩度、 v_i は画素 i の明度である。人が似ていると感じる色どうしは HSV 色空間上で近い位置に写像されると考えられるため、提案手法では色の類似度を把握するのに RGB の値ではなく HSV 色空間での座標を利用している。

4 HSV 色空間における座標と SegNet のラベルを利用した風景画像検索

ここでは、提案する HSV 色空間における座標と SegNet のラベルを利用した風景画像検索について説明する。提案手法では、画像の各画素の HSV 色空間における座標と SegNet の出力ラベルをリサイズして縮小したものを特徴量として使用する。画像に対応する特徴量をあらかじめ抽出しておき、画像と特徴量を関連付けてデータベースに保存する。検索キーとデータベース中の画像の特徴量を比較することで類似した画像を検索する。SegNet から出力されたラベルの情報だけでなく画像の HSV の情報も検索に利用することで、画像に含まれるオブジェクトの情報と色の情報も考慮した検索が行えるようにしている。

提案手法において、検索キーと画像 p の特徴量間の距離は SegNet のラベルが一致しなかった画素数と、検索キーと画像 p の同じ位置の画素の HSV 色空間における座標の差から求める。検索キーと画像 p の距離 $D^{(key,p)}$ は

$$\begin{aligned} D^{(key,p)} &= M^{(key,p)} + \frac{\alpha}{9} \left\{ \sum_{i=1}^N \left(x_i^{key} - x_i^{(p)} \right)^2 \right. \\ &\quad \left. + \left(y_i^{key} - y_i^{(p)} \right)^2 + \left(z_i^{key} - z_i^{(p)} \right)^2 \right\} \end{aligned} \quad (7)$$

で与えられる。ここで、 $M^{(key,p)}$ はキー画像と画像 p とで SegNet のラベルが一致しなかった画素数である。また、右辺の第 2 項は HSV 色空間における各画素の距離の 2 乗になっている。各画素において距離の 2 乗を取り得る最大値が 9 であるため、9 で割ることで正規化を行っている。 α は 2 つの項の重みづけを行うための係数である。また、 $x_i^{key}, y_i^{key}, z_i^{key}$ は検索キーの画素 i の HSV 色空間における座標、 $x_i^{(p)}, y_i^{(p)}, z_i^{(p)}$ は画像 p の画素 i の HSV 色空間における座標である。

検索を行う際には、この距離の値がしきい値以下の画像がすべて検索結果として出力されることになる。なお、係数 α の値や、リサイズ後の画像のサイズ、検索を行う際のしきい値などは実験を行った上で決定した。

5 計算機実験

文献 [3] において実験に用いられている 550 枚の画像を対象として検索実験を行った。提案手法において、領域のラベルと色情報を用いることで従来手法 [3] よりも高い精度で検索が行えることがわかった。

参考文献

- [1] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner : “Gradient-based learning applied to document recognition,” Proceedings of the IEEE, Vol.86, No.11, pp.2278–2324, 1998.
- [2] K. Simonyan and A. Zisserman : “Very deep convolutional networks for large-scale image recognition,” <https://arxiv.org/pdf/1409.1556.pdf>, (2021 年 7 月 18 日参照).
- [3] 矢森大貴, 長名優子 : “畳み込みニューラルネットワークを用いた風景画像を対象とした類似画像検索,” 情報処理学会第 83 回全国大会, 2021.
- [4] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, Jiaya Jia : Pyramid Scene Parsing Network <https://arxiv.org/abs/1612.01105>, (2022 年 1 月 5 日参照).
- [5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun : <https://arxiv.org/abs/1512.03385>, (2022 年 1 月 5 日参照).