

人の視覚特性を考慮した文字領域抽出

日大生産工(院) ○飛田 智史
日大生産工 松田 聖

1 まえがき

近年、画像処理の分野で人間の視覚をコンピュータで模擬しようとする研究が盛んに行われている。この研究は近い将来、ロボットヴィジョンや翻訳システム、画像検索への応用が考えられている。

文字抽出処理の対象は、大きく二つに分けられる。一つは文章画像を対象としたもので、スキャナーなどに応用されている。もう一つが情景画像を対象とした文字抽出である。情景画像を対象とした文字抽出は、多くの手法が提案されているが、二値化[1]や連結成分分析[2]、ニューラルネットワークによる学習[3]、パターンマッチングをもとにしたもの[4]がほとんどである。こうしたアプローチは、複雑な背景中に文字が存在したり、文字が一行に並んでいることや、学習による多くのデータが必要であったり、縦書き、横書き、対象フォントの制約などいくつかの問題点も報告されている。

そこで本研究では、文字抽出を行う背景画像の看板や文字部分は人間が作成したものであるという点に着目し、視認性と誘目性という人間の視覚特性を考慮した文字領域抽出アルゴリズムを提案する。

2 人の視覚特性

人間には区別が付きやすかったり、見えやすい色が存在する。これを色の視認性という。色の視認性では、背景との色の明度差がもっとも重要で、次に彩度差、色相差が大きいほど視認性が高くなるといわれている[5]。

また、人間の目につきやすいか、目に付きにくいかの性質を現すものを色の誘目性という。誘目性は動いている、止まっているなどによる要因が大きいですが、色に関しては一般的に無彩色よりも有彩色、青や緑よりも赤や黄色のほうが誘目性が高いとされている。

3 処理概要

処理概要をFig.1に示す。デジタルカメラで撮影した情景画像を入力として、均等色空間である $L^*a^*b^*$ 表色系に変換を行う。

$L^*a^*b^*$ 表色系は、色を見たときに、同じ色違いに見える色同士の距離を均等にしてある色空間のことである。 L^* 、 a^* 、 b^* の値は次の非線形変換で得られる。

$$\begin{aligned} L^* &= 116(Y/Y_0)^{1/3} - 16 \\ &\quad \text{if } (Y/Y_0) > 0.008856 \\ L^* &= 903.29(Y/Y_0) \\ &\quad \text{else,} \\ a^* &= 500[(X/X_0)^{1/3} - (Y/Y_0)^{1/3}], \\ b^* &= 200[(Y/Y_0)^{1/3} - (Z/Z_0)^{1/3}], \end{aligned} \quad (1)$$

ここで、定数 X_0, Y_0, Z_0 は光源の X, Y, Z 値である。 X, Y, Z 値は R, G, B 値を使って次式で定義した。

$$\begin{aligned} X &= 0.41R + 0.36G + 0.18B, \\ Y &= 0.21R + 0.72G + 0.07B, \\ Z &= 0.02R + 0.12G + 0.95B. \end{aligned} \quad (2)$$

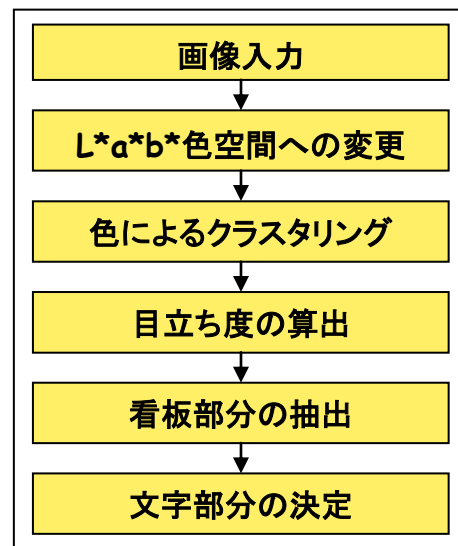


Fig.1 処理概要

ここで、 L^* は輝度、 a^* は緑と赤を両端とした色相や彩度の変化、 b^* は青から黄色の変化を示す。

$L^*a^*b^*$ 色空間変更後、色によるクラスタリングを行う。クラスタリングは非階層型クラスタリング手法であるK平均アルゴリズムを用い、2点間の距離尺度としては、 $L^*a^*b^*$ 空間におけるユークリッド距離を用いた。また、最適なクラスタ数 k の決定は、一般的に難しい問題であるが、様々な研究による論文から $k=6$ とした。

その後、画像中の目立ち度を算出し、看板らしい部分を抽出する。最終的に看板と思われる領域の中で目立ち度の高い領域を文字領域とし、二値化して文字領域であるか否かを人間の目によって判定する。

4 目立ち度の算出

画像中の存在感、つまり目立ち度を計算する。目立ち度は、色彩科学の分野から以下の3つの項目が重要とされている[6]。

- ・ 色領域の大きさ
- ・ 誘目性の高さ
- ・ 視認性の高さ

1番目の「色領域の大きさ」は、画像中に占める領域の大きさである。これはクラスタリングした後の同じ色の要素の和で算出する。

$$C_1 = \sum x(L^{*i}, a^{*i}, b^{*i}) \quad (3)$$

2番目の「誘目性の高さ」とは、色自体が持つ人間の視覚への目につきやすさである。これは $L^*a^*b^*$ 色空間の原点を変更し、 L^* と a^* と b^* の二乗平均で算出する。

$$C_2 = \sqrt{L^{*2} + (a^* + 60)^2 + (b^* + 60)^2} \quad (4)$$

3番目の「視認性の高さ」とは、コントラスト感の高さであり、文字色と地色の明度差の大きさである。これは輝度の差であるので、 L^* 軸、 a^* 軸、 b^* 軸の差をそれぞれ二乗したものの和の平方根で算出する。

$$C_3 = \sqrt{\Delta L^{*2} + \Delta a^{*2} + \Delta b^{*2}} \quad (5)$$

以上3つの数値から、色領域の大きさを2、誘目性の高さを5、コントラスト感の高さを4で重み付けし二乗平均を目立ち度とする。

$$L = \sqrt{(2C_1)^2 + (5C_2)^2 + (4C_3)^2} \quad (6)$$

5 看板領域の抽出

画像中の文字が書かれている看板部分の抽出を行う。看板部分は画像中にそれなりの面積をしめると考えられる。また、一般に看板は

長方形、円形、三角形等の単純な形をしていると考えられる。この特徴を利用して、クラスタリングによって分割された領域の面積と画像全体を比べて、明らかに小さい部分は削除し、単純な形をしている領域を看板候補領域として抽出する。そこで円形度 C を用いる。

$$C = (4\pi S) / U^2 \quad (7)$$

ここで、 S は領域の面積、 U は周囲長である。 C の範囲は $0 < C \leq 1$ であり、円が1.0、正方形が0.79、正三角形が0.6である。形が複雑になると C の値は小さくなるため、 C の値が小さいものは看板領域ではないとする。

5 文字部分の決定

文字領域の決定は看板領域の中の目立ち度の高い色の領域とする。また、注目領域、つまり看板部分の中で色差が小さい領域が複数ある場合は、色を結合する処理を行う。これはクラスタ数固定の問題を解決するために行う。

最終的に文字領域と決定した領域と看板の領域を二値化する処理を行い、看板部分だけを抜き出すことで文字領域が正常に抽出できたか否かを人間の目によって判断する。

「参考文献」

- [1] 劉詠梅、山村毅、大西昇、杉江昇、シーン内の文字領域の抽出について、電子情報通信学会論文誌、(1998)、pp.641～650
- [2] 松尾賢一、上田勝彦、梅田三千雄、適応しきい値法を用いた情景画像からの看板文字列領域の抽出、電子情報通信学会論文誌 (1997)、pp.1617～1626.
- [3] 岩片智、味岡義明、萩原将文、局所並列処理による情景画像からの文字領域抽出アルゴリズム、電学論 (2004)、pp.959～965
- [4] 安居院猛、関根証明、佐野元昭、C言語による画像処理演習 昭晃堂
- [5] 大田登、色彩工学、東京電機大学出版局
- [6] 日本色彩研究所編、カラーコーディネーターのための色彩科学入門 日本色研事業株式会社