

ポスター画像中のイベント情報抽出システム

柳井美保※, 河野恭之※ (※関西学院大学)
 {dov97462, kono}@kwansei.ac.jp

目的

告知に用いられるポスター画像中から
 イベント情報を抽出する手法を提案



保管したポスター画像をOCRエンジンに読み込ませ、テキストデータへ変換し、電子カレンダーに記載できるようにする

背景

ポスターのイベント情報を画像情報として保管する人が増加している。

例

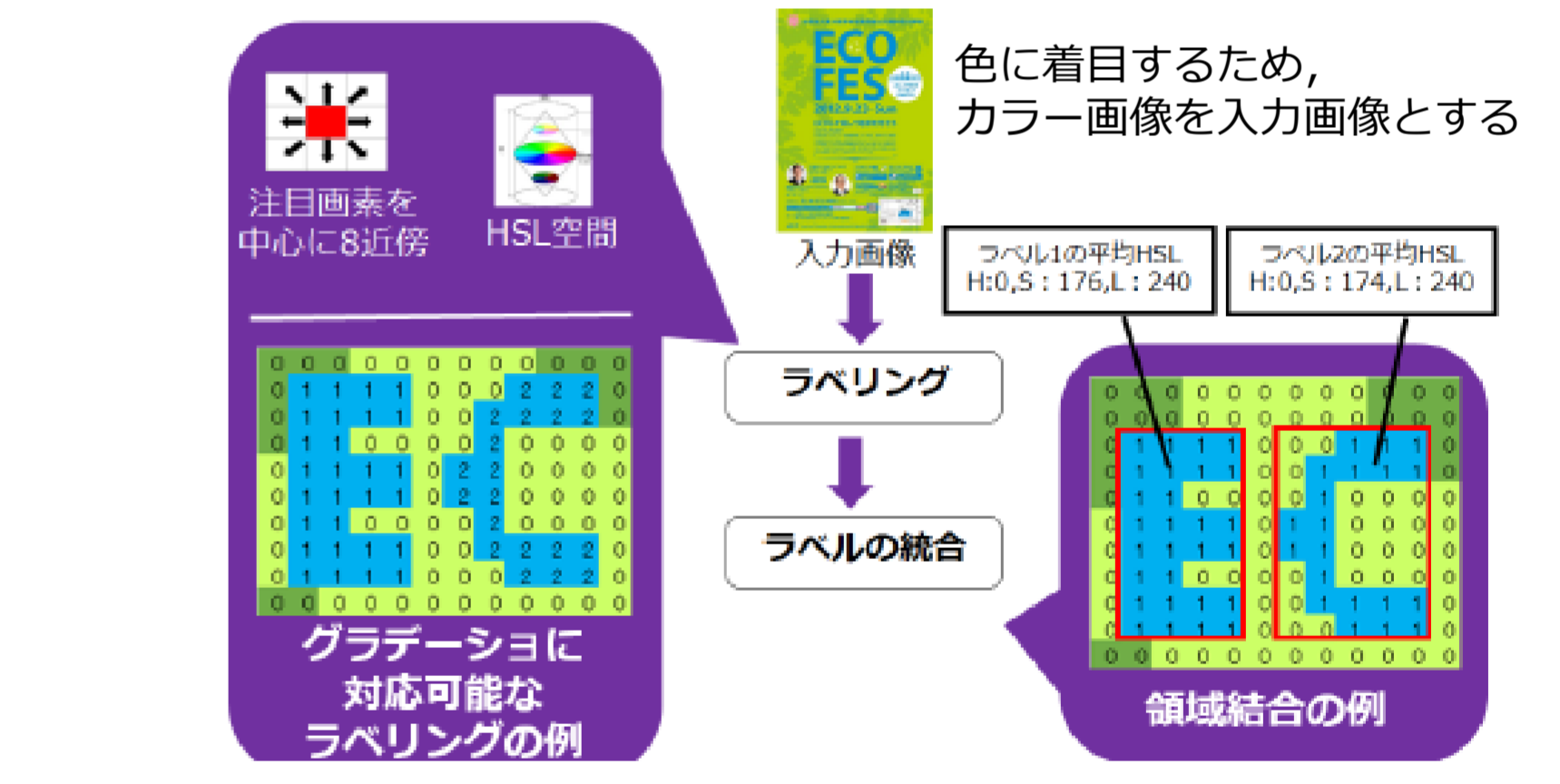


保管した大量の画像情報から目当ての情報を探すことは手間がかかる。

提案手法の概要



Key point : 文字に近い領域に絞る



ラベリング

注目画素を中心に8近傍に対してHSL空間上のユークリッド距離を求める。この距離がしきい値以下であれば同じラベルを振り、しきい値以上であればラベルを振らない。

ラベルの結合

1つの文字列に1つのラベルを振り分けるため結合を行う。

イベント名 : ECOFES
 日時 : 2012.9.23

イベント名の選定

1文字のフォントサイズが最も大きい
 1文字のフォントの大きさを求める計算

$$\frac{\text{文字領域を包含する面積}}{\text{文字数}}$$

日時の選定

- ・ イベント名以外の矩形
- ・ 数字が2桁以上含まれている
- ・ ホームページアドレスや電話番号、
- ・ 値段、定員、住所が含まれてない