

平成23年度 情報工学コース卒業研究報告要旨

大西 研究室	氏 名	星 原 光 貴
卒業研究題目	新聞記事スキャン時に生じる裏写り除去及び部分画像の結合手法	

背景と目的

新聞紙は劣化しやすく、かさばるため、スキャナを用いて電子データとして保存する人が増えている。しかしながら、一般家庭において新聞記事をスキャンすることを考えた場合、次のような問題が生じる。

まず、一般家庭用スキャナはA4サイズまでしか対応していないことが多く、新聞紙のような大きなサイズを一回でスキャンすることは不可能である。次に、新聞記事は薄い紙に印刷されるため、スキャン時に裏写りが生じる。最後に、新聞紙の下地の灰色は、電子データとして残す上で見栄えを悪くする原因となる。

そこで、本研究では、スキャナを用いて得られた新聞記事の部分画像を結合し、新聞紙の下地色及び裏写りを除去するシステムを実現する手法を提案する。

裏写り除去手法

既存の手法として、画像を二値化することで裏写りを除去する手法があるが、新聞記事の色調情報を失ってしまうため、カラー記事や写真を含む記事には適さない。

そこで、本研究では、コントラストを強調することで、新聞記事の色調情報を残したまま裏写りを除去する手法を提案する。また、本手法によって新聞紙の下地及び裏写りが除去できる一方で、記事中のオブジェクトについては消滅しないことを示した。

部分画像の結合手法

部分画像の結合処理について説明する。まず前処理として、文字の並び情報を利用した傾き補正を行い、傾きを0.125度以下に抑える。

次に、二段階テンプレートマッチングを行う。第一段階目のテンプレートマッチングでは、画像中の2箇所からテンプレート画像を取得し、テンプレートマッチングを行う(図1)。その結果、取得元の2箇所を結ぶ線分とマッチング先の2箇所を結ぶ線分が得られる(図1の①及び②)。そして、当該2つの線分の傾きから「前処理で補正しきれなかった細かい傾き」を補正すると共に、当該2つの線分の長さの比率から「スキャン時に生じる画像の歪み」の補正を行う。その後、第二段階目のテンプレートマッチングにおいて、画像の結合位置を探索し結合する。

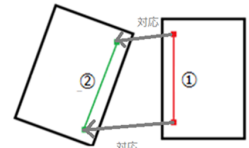


図1 第一段階目

実験と結果

新聞記事30記事(画像60枚)に対し、実際に結合処理を行ったところ、30記事中、29記事について結合の境目付近に存在する全ての文字が読み取れる状態で結合することができ、本手法の有効性が確認できた。図2は、見開きの新聞記事を10回に分けてスキャンしたものを結合したものである。また、新聞記事30記事に対し、裏写り除去処理を行ったところ、十分な裏写り除去の効果が認められた。図3及び図4に裏写り除去の例を示す。

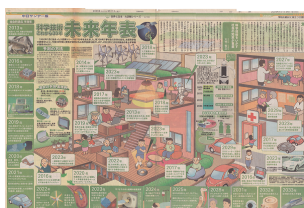


図2 見開き記事の結合結果
朝日新聞(2005年9月11日)



図3 裏写り除去処理前
日本経済新聞(2011年12月19日)

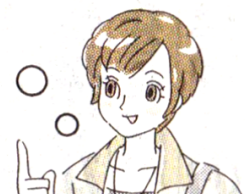


図4 裏写り除去処理後
日本経済新聞(2011年12月19日)