

# 動画像中のテキスト認識を目的とした文字形状に基づくビデオモザイク手法

宮崎 洋光\* 内田 誠一\*\* 迫江 博昭\*\*

\*九州大学大学院 システム情報科学府 \*\*九州大学大学院 システム情報科学研究院

## 1 はじめに

カメラによる画像中のテキスト認識、理解において、動画像の利用が検討されている。動画像を用いることの利点として、静止画像では画角の制約により撮影できない長いテキストであっても撮影できることが挙げられる。

動画像を用いた場合、テキストは複数のフレームに断片的に撮影される。加えて、撮影の際の手ぶれにより回転、拡大縮小、上下移動、カメラ移動速度などの変動が生じる。従って、テキスト認識を行うためには、手ぶれにより生じる変動を何らかの方法で補償しながら、フレーム統合を行う必要がある。本報告では、フレーム統合処理とテキスト認識処理の2つの処理を同時に一括して行う手法を提案する。本報告の詳細に関しては [1],[2] を参照されたい。

## 2 手法の概要

本手法では、手持ちカメラを左から右へと移動させながらテキストを撮影したビデオフレーム列を対象とする。解くべき問題の簡単化のために、撮影されたビデオフレーム列について、各フレームの中心より幅  $W$  画素、高さ  $H$  画素の矩形領域（以下、スリット）を考える（図 1 (a)）。

フレームに生じる変動は、回転、拡大縮小、上下移動、カメラ移動速度を想定している。まず、カメラ移動速度の変動のみを含む場合について説明を行う。この場合は、図 2 に示すように入力フレームを横軸に、各標準文字パターンを縦軸にとった平面中の最適経路問題として表現され、その最適化は DP アルゴリズムを用いて効率的に求めることが可能である。

具体的には、初期フレームから最終フレームまでのスリットと標準文字パターンの各領域との局所距離および最小累積距離を順次求めて行く。同図の点  $(t, j)$  において最小値選択を行う際には、先行可能な点  $(t-1, j')$  ( $j' \in \{j, j-1, j-2\}$ ) より最小値を与える。この処理によりカメラ移動速度の変動、すなわち各フレームの水平方向の変動を補償することが可能となる。最終的に最小累積距離を与える経路をバックトラックすることによりテキスト認識の結果を求める。

次に、カメラ移動速度の変動に加え、回転、拡大縮小、上下移動の変動が含まれる場合について説明する。この場合の着目点は、フレームに生じた変動に合わせてスリット形状を制御することである。具体的には、図 1(b) に示すように、回転を  $r$ 、拡大縮小を  $s$ 、上下移動を  $\delta$  で表し、それぞれ画素単位で制御する。スリット形状の制御が最適に行われた場合、前述したカメラ移動速度の変動のみの最適マッチング問題と同様に考えることができる。しかし、実際にはスリット形状の最適な制御は未知であるので、各スリットについて全ての形状制御を試みながら最適マッチングを求める必要がある。

## 3 実験

本実験では、(1) カメラ移動速度変動のみを含む、(2) 全ての変動を含む、(3) 個別に変動を含む、それぞれの場合について認識実験を行い手法の有効性を検討する。実験資料として、紙面に印刷された英数字をカメラ移動速度  $0 \sim 2$  pixel/frame で撮影した動画像データを用いる。

実験結果を図 3 に示す。横軸はフレームに生じる  $r, s, \delta$  の最大変動量である。(1) については、認識率は約 95% であった。すなわち、速度変動がほぼ補償されており、本手法の有効性が確認できた。(2) については、変動が大きくなるに従い認識率は低下している。さらに (3) については著しく低下している。原因として、特徴量として2値化画像を用いたためにマッチングの際に大きな量子化誤差が生じ、上手く補償ができなかったことが考えられる。今後、さらに詳しく考察し改善を行う必要がある。

## 4 まとめ

本報告では、動画像中のテキスト認識を目的として、手ぶれ変動の補償、フレームの統合ならびに認識を同時に一括して実行する手法を提案した。

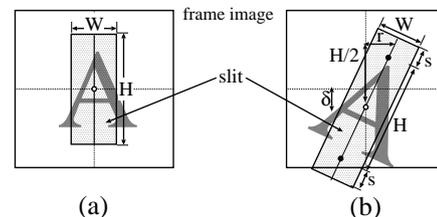


図 1: (a) スリット ( $r = s = \delta = 0$ )、および (b) 形状制御されたスリット

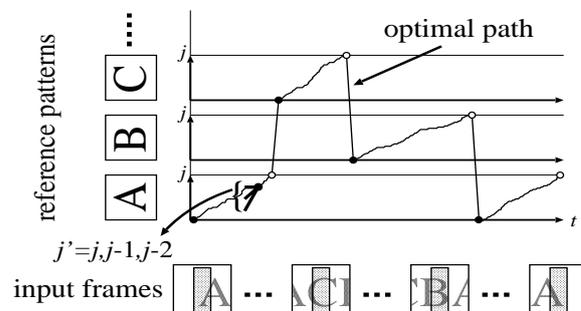


図 2: カメラ移動速度変動のみを考慮した DP マッチング

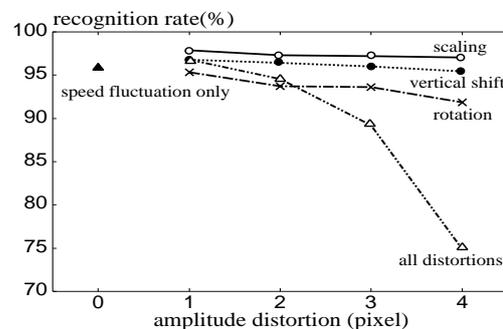


図 3: 実験結果

## 参考文献

- [1] 宮崎, 内田, 迫江, “動画像中のテキスト認識を目的としたビデオモザイク手法,” 信学技報, PRMU2004-248, March, 2005.
- [2] 宮崎, 内田, 迫江, “文字形状に基づく認識駆動型ビデオモザイク手法,” MIRU2005, OS7B-35, July, 2005.