

競技映像における文字読み取りの検討

宮原 景泰 依田 文夫
三菱電機(株) 情報技術総合研究所

1. はじめに

競技映像では、視聴者に競技状況を知らせるための文字がスーパーインポーズされるため、この文字を読み取ることで、所望の状況の映像を自動で取得・検索することが可能となる。しかし、一般的な映像検索を目的とした従来のスーパーインポーズ文字の読み取り手法^[1]は、OCRと同じように文字切り出し→認識のシーケンスを探っており、屋外競技の映像のようにカメラ視野内の情景変動が激しい場合や、特殊フォントで情景中にスーパーインポーズされるパターンへの適用は難しい。一方、競技映像の表示パターンは毎回決まっているので、あらかじめパターンを登録しておくことは容易であり、事前登録を行わない汎用的な手法よりも、簡単な登録操作で安定した読み取りを実現するような手法が望ましいと考えられる。

そこで、競輪競技の映像を対象として、パターンの事前登録を前提としたスーパーインポーズ文字の読み取り手法について検討を行ったので、本稿では、この検討内容を報告する。

2. 検討内容

一般的な文字読み取り手法を競技映像に適用した場合の大きな課題として、文字切り出しの難しさが上げられる。図1はこの例であり、情景中に文字が中抜きフォントで表示され、さらに文字パターンと情景中の白線とが接触しているケースである。所望の映像部分を検索するという観点では、必ずしも表示された全文字を認識する必要はなく、検索対象に対応した特定のパターン（文字パターン）が画像中に存在するか否かが判定できれば良いため、今回検討する手法では文字切り出しは実施せず、画像から直接文字パターンを認識するアプローチを探ることにした。

画像から特定のパターンを直接検出・認識する手法としては、テンプレートマッチングが代表的である。

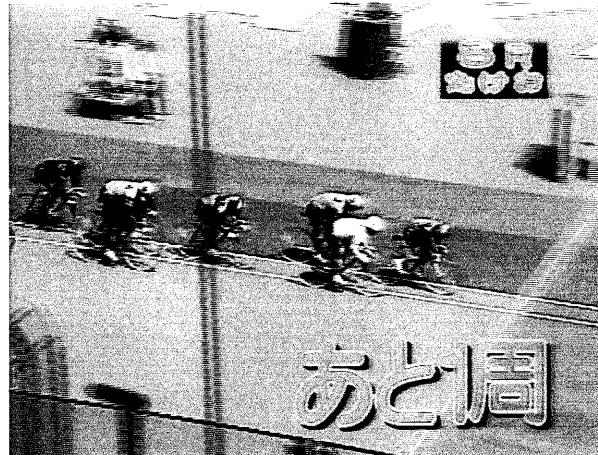


図1 競技映像中のスーパーインポーズ文字例

しかし、一般的なテンプレートマッチング手法をそのまま適用した場合、(1)部分的に文字パターンに類似した情景領域がスーパーインポーズ位置に来ると誤認識する恐れがある、(2)類似パターンの判別能力が十分ではない、という課題がある。このため、文字パターンを複数の部分領域に分割して部分領域毎のマッチング結果で文字であるか否かを判定すると共に、異なるパターンの差異を強調するよう画素単位に重み付けしてマッチングを行う構成とした。

図2に本手法の基本フローを示し、以下、各処理の内容を説明する。

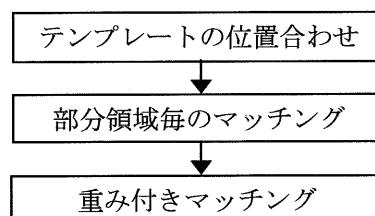


図2 検討手法の基本フロー

2. 1 テンプレートの位置合わせ

認識対象はスーパーインポーズされた文字パターンであり、出現位置はほぼ決まっているが、わずかに位置ずれする場合もあるため、最初にテンプレートの位置合わせを行う。照合位置をずらしながら式1に示す従来のテンプレートマッチング処理（ただし、文字

パターンを構成する有効画素のみ対象とする)を行い、相違度(残差)が最小となる位置を求め、この位置を以降のマッチングにおける照合位置とする。

$$\text{相違度} = \frac{\sum_{x,y} |I(x+u, y+v) - T(x, y)|}{N} \quad \dots \text{(式1)}$$

但し、Iは入力画像、Tはテンプレート画像
Nはマッチング対象画素の総数

2.2 部分領域毎のマッチング

あらかじめテンプレートを輝度値に従って2種類の部分領域(黒っぽい領域と白っぽい領域)に分けておき、この領域別に式1のマッチングを行って、得られた相違度の最大値(評価値1)が閾値を越える場合は棄却と判定する。スーパーインポーズされる文字パターンには一定のコントラストがあり、これを利用してことで、任意の情景と文字パターンとの判別を強化できる。

2.3 重み付きマッチング

画素単位の重み情報Wを用いた式2のマッチングを行って相違度を求め、相違度最小のテンプレートと2番目に小さいテンプレートとの相違度差(評価値2)を算出する。この値が閾値未満の場合は棄却と判定し、閾値以上ならば相違度最小のテンプレートを認識結果とする。

式2の重み情報Wは、異なるパターンの差異を強調することを目的とし、パターン間の差分画像を基に、式3に示す学習動作を繰り返してあらかじめ作成しておくものである。なお、差分画像は、対象とする両パターンで共に有効画素でなければ0、片側のみが有効画素なら定数としておく。

$$\text{相違度} = \frac{\sum_{x,y} \{|I(x+u, y+v) - T(x, y)| \times W(x, y)\}}{N} \quad \dots \text{(式2)}$$

$$\left. \begin{array}{l} \text{if } Aji \cdot Wi(t) \leq TH_L \\ \quad Wi(t+1) = Wi(t) + \alpha(t) \left\{ \frac{Aji}{|Aji|} - Wi(t) \right\} \\ \text{else} \\ \quad Wi(t+1) = Wi(t) \end{array} \right\} \dots \text{(式3)}$$

但し、Wi(t)は時刻tにおけるテンプレートiの重み情報
Ajiはテンプレートiとjの差分画像
 $\alpha(t)$ は単調減少関数であり、TH_Lは定数
なお、Wiは更新される毎に2乗和一定に正規化する。

3. 実験

図1に示した映像のレース番号(図1では右上の8)を対象として認識実験を行った。本映像は8レースの後半から9レースの始めまでであり、レース間では文字は表示されていない。従来法(式1)の結果も含め、本実験の結果を図3と図4に示す。従来法と本手法共に誤認識することはなかったが、本手法は従来法に比べ、文字が表示されていない時の評価値1がより大きく、また文字が表示されている時の評価値2もより大きいことから、認識の安定性が向上していると言える。

評価値1

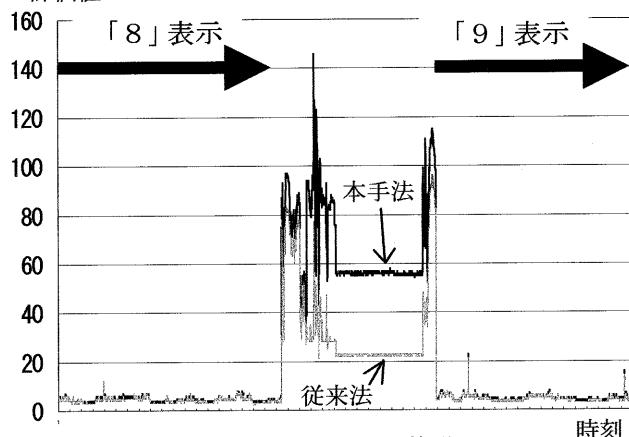


図3 評価値1の推移

評価値2

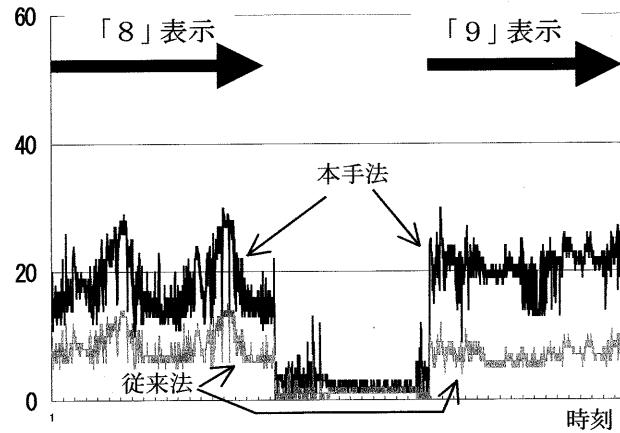


図4 評価値2の推移

4. まとめと今後の課題

競技映像向けの文字読み取り手法を検討し、競輪映像で検討手法の効果を確認した。今後はより多くの映像について検証を進める予定である。

参考文献

- [1]森ほか：“画質劣化にロバストな映像中テロップ文字認識”，信学技報 PRMU98-154 (1998)