

移動するテロップの自動認識

Recognition Method for Moving Text in Video

五十川 賢造[†] 三田 雄志[†] 井田 孝[†] 堀 修[†]
 Kenzo Isogawa Takeshi Mita Takashi Ida Osamu Hori

1 はじめに

番組検索のためのキーとしてテロップが注目されている [1]。特に、番組の最後に表示される右から左もしくは下から上に移動するテロップ (図 1) には出演者名や配役等が EPG (電子番組表) より多く含まれている。これらを映像から自動認識してテキストデータにできれば有用である。

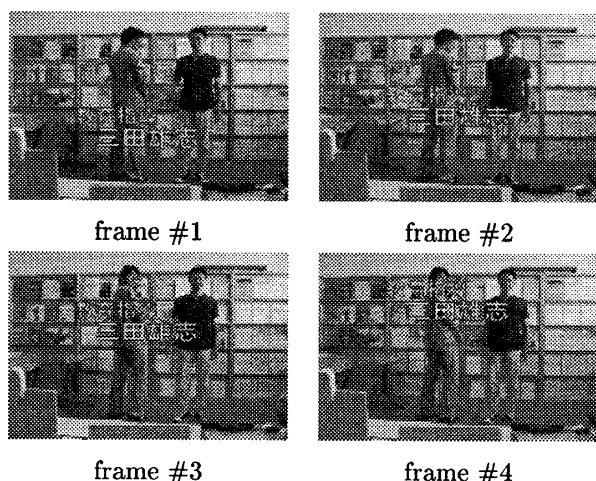


図 1: 下から上に移動するテロップの例

テロップの認識においてはまず、テロップの画素に 1 を、それ以外の画素に 0 を割り当てた二値画像 (テロップ二値画像) を生成する。次に、それを OCR (Optical Character Reader) で認識してテキストデータを得る。このとき、テロップは番組によって異なるストローク幅や輝度を持ち、また図 1 のように複雑なテクスチャを含む背景の上に表示されると、固定の閾値を用いた 2 値化のような単純な処理では、OCR で認識可能なテロップ二値画像を得ることはできない。それに対し筆者らは複数枚のフレーム画像を用い、高い輝度を持ちなおかつ輝

[†](株) 東芝 研究開発センター マルチメディアラボラトリー, Multimedia Laboratory, Corporate Research & Development Center, Toshiba Corporation

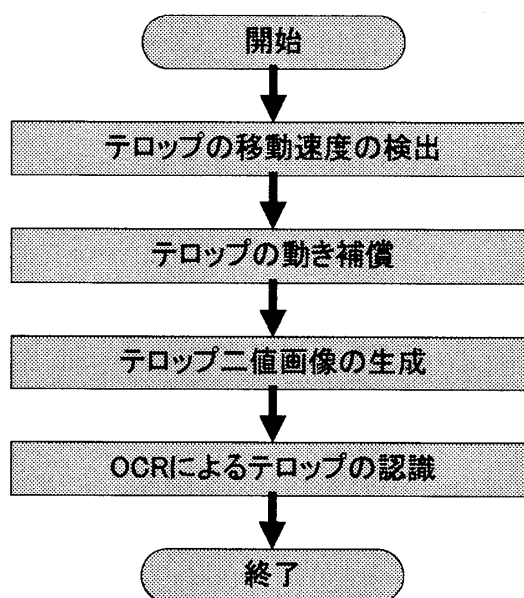


図 2: 提案手法によるテロップの認識の流れ

度の時間変化が少ない画素を抽出してテロップ二値画像を生成する手法 [2] を提案した。この手法はニュースの見出しのように移動しないテロップを認識できるが、移動する場合、あるフレームでテロップの画素であってもその前後のフレームでは輝度に変化するため、それを抽出できない。

そこで本稿では移動するテロップを自動認識するため、まずテロップの移動速度を検出し、次に求めた速度に基づいてテロップの動き補償を行った後に、文献 [2] の手法を用いてテロップ二値画像を得る手法を提案する (図 2)。この手法においては、テロップの移動速度を高精度に検出する必要がある。そのために今回、速度を時空間スライス画像におけるテロップの軌跡の傾きとして検出した。

次章でまず提案手法について述べる。次に実験により提案手法の速度の検出精度とテロップ二値画像の認識精度を評価した結果を示す。

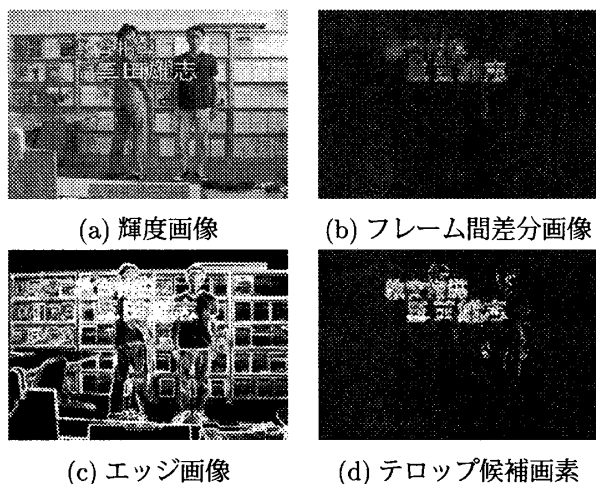


図3: テロップ候補画素の抽出過程。フレーム間差分値 (b) が大きくかつエッジ強度 (c) の大きい画素をテロップ候補画素 (d) とする。

2 提案手法

テロップは人が読みやすいように、(a) 輝度は一様で背景より高い、(b) 輝度は時間変化しない、(c) 一定時間以上表示される、(d) 画面内を一定速度で移動する、(e) 画面内の全てのテロップが同じ速度で移動する、という特徴を持つものが多い。本研究では上記の5つの性質を仮定する。以後、図2の流れに従い提案手法を説明する。

2.1 テロップの移動速度の検出

移動速度の検出は、(1) テロップ候補画素の抽出、(2) 時空間スライス画像の作成、(3) テロップの軌跡の検出、(4) 軌跡からの速度の算出の4ステップで行う。

2.1.1 テロップ候補画素の抽出

テロップの輪郭の画素は、前に述べた性質 (a) から高いエッジ強度を持つ。また性質 (a), (d) から輝度値の時間変化量が大きくなる。そこで、フレーム間差分値 (図3(b)) が一定以上であり、エッジ強度 (図3(c)) が一定以上である画素をテロップ候補画素 (図3(d)) とする。

この段階では例えば、図3の「田」の縦線は輝度の時間変化がないため検出に失敗している。また、人物が動いているため、テロップではないがその輪郭も抽出される。このような誤検出があっても、以後の処理により速度の検出が可能となる。

2.1.2 時空間スライス画像の作成

x (画面横方向)、 y (画面縦方向)、 t (時間方向) の3次元空間にテロップ候補画素を配置し (図4(a))、それを x が一定の平面で切断して、時空間スライス画像 (図4(b))

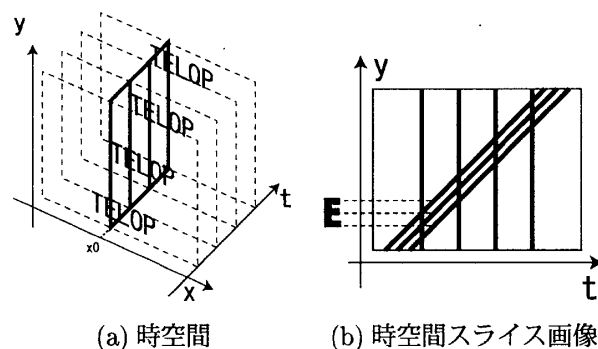


図4: 時空間におけるテロップ。 x, y, t の3次元空間 (a) にテロップ候補画素を配置し面 $x = x_0$ で切断する。得られる時空間スライス画像 (b) 上ではテロップ候補画素が直線上に並ぶ。この直線がテロップの軌跡であり、その傾きが移動速度である。

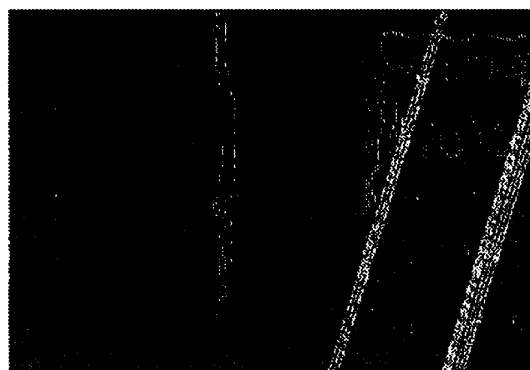


図5: 実際の時空間スライス画像。座標軸は図4(b)と同じ。右側にテロップの軌跡が集まってできた帯が2本表示されている。

を得る。時空間スライス画像上では、テロップの軌跡がテロップ候補画素がなす直線として現れる。この直線の傾きがテロップの移動速度である。

実際の時空間スライス画像 (図5) では直線が欠けたり、テロップの輪郭以外の画素も2.1.1で述べた誤検出により現れる。

2.1.3 テロップの軌跡の検出

時空間スライス画像からテロップの軌跡を検出する。まず、検出処理を高速化するためテロップの軌跡を損なわないように画素を間引く。性質 (a), (b) からテロップの軌跡上に存在する画素では、フレーム画像上で輝度勾配の方向が時間変化しない。そこで、フレーム画像上で真上方向の輝度勾配を持つ画素のみを抽出した時空間スライス画像を生成する。次に RHT (Randomized Hough

Transform)[3]を用いて、テロップの軌跡を直線として検出する。同様に、真下方向の輝度勾配を持つ画素がなす直線も検出する。

RHTは以下に述べる3ステップから成る。まず、画面内からランダムにテロップ候補画素を2点サンプリングし両点を通る直線のパラメタを求める操作を、所定の回数行う。次に、得られたパラメタ群の中から最も出現頻度が高いものを選ぶ。最後に、選ばれたパラメタに対応する直線上に実際に存在するテロップ候補画素の数を数え、それが所定の閾値 l_t を超えた場合はその直線をテロップの軌跡として採択する。上記3ステップを、どの直線にも属さないテロップ候補画素の数が一定値以下になる、もしくは繰り返し回数が事前に決めた上限値に達するまで繰り返す。

RHTはハフ変換と同様に直線の欠損やノイズにロバストである。また画面上のテロップ候補画素の位置から直線の傾きを求めるため、投票空間の解像度設定が不要であり直線のパラメタの検出精度が高まる。また l_t の値を調整することで、性質(c)を満たすもののみを抽出できる。テロップの軌跡の傾きを検出することは複数のフレームの情報を統合することであり、速度を高精度に検出できる。

2.1.4 軌跡からの速度の算出

図5のように時空間スライス画像上にはテロップの軌跡が複数現れる。それぞれがRHTにより検出されるが、性質(e)が満たされればテロップの軌跡は全て同じ傾きを持つ。しかし実際には検出される移動速度には多少のばらつきがあるので、それらの重み付け平均を最終的なテロップの移動速度とする。具体的には、時空間スライス画像上に直線が N 本検出され、 $j(j=0, \dots, N-1)$ 番目に検出された直線が速度 v_j に相当する傾きを持ち、その直線上のテロップ候補画素の数が l_j であるとき、テロップの移動速度 v を

$$v = \frac{\sum_{j=0}^{N-1} \frac{l_j v_j}{l_{max}(v_j)}}{\sum_{j=0}^{N-1} \frac{l_j}{l_{max}(v_j)}} \quad (1)$$

とする。 $l_{max}(v_j)$ は速度 v_j のテロップの軌跡上に本来存在すべきテロップ候補画素の数であり、入力画像の高さ h とフレーム周期 Δt から

$$l_{max}(v_j) = \frac{h}{v_j * \Delta t} \quad (2)$$

である。

2.2 テロップの動き補償

図6に示すように複数枚のフレーム画像をテロップの移動速度に応じてずらし、テロップの位置を合わせてか

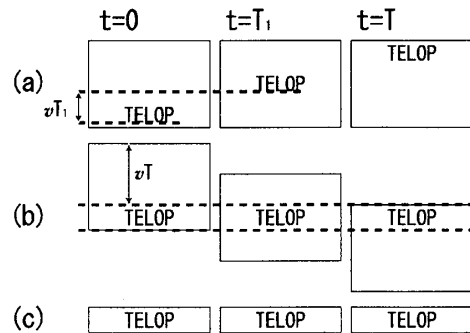


図6: テロップの動き補償。(a) フレーム画像を取得し、(b) テロップの移動速度 v に応じてずらし、(c) トリミングしてテロップの位置が補償された動画像を得る。



図7: 図1の動画像に対しテロップの動き補償を行い、画素毎の輝度値の時間平均をとったもの。テロップ以外の部分はブレて表示される。

ら輝度の時間平均を取る。図1の4枚のフレーム画像のテロップの動きを補償した結果を図7に示す。

なお提案手法では、新たなフレーム画像が入力される毎に取得済みのフレーム画像のうち最も古いものを廃棄して、テロップ二値画像の生成とOCRによるテロップの認識を行う。そのため、動き補償に利用する映像区間長より長い時間表示されているテロップに対しては、複数の認識結果が得られる。

2.3 テロップ二値画像の生成

動き補償によってテロップの画面内での位置が固定されるので、文献[2]の手法を用いてテロップ二値画像(図8)を得る。

この時、例えば図1の本棚や人は動き補償により、下方向に移動している(図7)ため、輝度が時間変化し背景と判断される。テロップと同じ速度で画面内を移動するオブジェクトはほとんどないため、テロップ以外の画素をテロップと誤認識することは少ない。

2.4 OCRによるテロップの認識

テロップ二値画像をOCRに入力し、テロップを認識する。本研究で用いたOCRは、認識結果のテキストデー

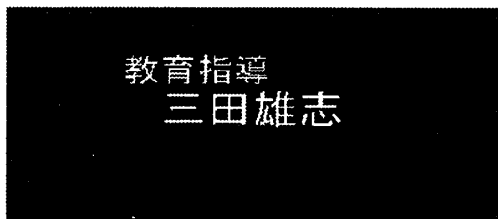


図 8: 図7から得られたテロップ二値画像。これをOCRに入力する。

表 1: テスト映像の特徴

	フォント	背景
映像 1	明朝体、15×15pixel	娯楽施設内部、人物
映像 2	明朝体、15×15pixel	芝生、樹木、人物
映像 3	明朝体、20×20pixel	建物外壁、事務所内、人物
映像 4	ゴシック体、15×15pixel	室内、田園風景、駅の雑踏、人物

タ、画像と文字の類似度、文字の外接矩形を出力する。

提案手法では 2.2 で述べたとおり同一のテロップに対し複数回の認識処理が行われるので、その結果を統合する必要がある。そのため、得られた外接矩形と速度の検出結果からその重なりが大きい物を同一のテロップとし、類似度が最も大きいテキストデータをそのテロップの認識結果とする。

3 実験

提案手法による移動速度の検出精度と認識精度を検証した。実験には 360×240 画素、30fps の 4 種類のテレビ映像、合計 232 秒を使用した (表 1)。正解データとして全てのテロップの外接矩形とテキストデータを手作業で作成した。

表 2 にテロップの平均移動速度を示す。速度の検出誤差は 0.4pixel/sec 以下であった。今回、動き補償は長さ 0.9 秒の映像区間で行ったので、テロップの動き補償の誤差は 0.5pixel 以下であり、十分な精度と言える。

表 3 に、文字総数に対する正しく認識された文字数の割合 (文字再現率) と、検出された文字数に対する正しく検出された文字数の割合 (文字適合率) を示す。再現率が高いことはテロップの見落としが少ないことを示し、適合率が高いことは誤検出が少ないことを示す。いずれも 100%に近いほど精度が高いといえる。文字適合率が 95.1%において 54.6%の文字を認識できた。線が太くて大きいフォントのテロップが多用された映像 3 では認識精度が高く、線が細い明朝体を用いた映像 2 では文字の横線の抽出に失敗する例が多く見られた。明朝体の横線

表 2: テロップの移動速度の検出結果

	映像区間長 (秒)	正解速度 (pixel/sec)	検出速度 (pixel/sec)	検出誤差 (pixel/sec)
映像 1	21	42.7	42.6	0.1
映像 2	26	56.1	56.0	0.1
映像 3	104	97.8	97.6	0.2
映像 4	81	74.6	74.2	0.4

表 3: テロップの認識結果

	文字総数	文字再現率 (%)	文字適合率 (%)
映像 1	160	46.8	98.1
映像 2	184	25.1	90.5
映像 3	630	72.0	96.2
映像 4	652	48.1	94.5
総計	1626	54.6	95.1

は周囲の色のにじみなどで暗くなる傾向があり、テロップの色が一様であるという仮定が成り立たないために失敗した。

4 おわりに

本稿では映像からテロップの移動速度を検出し動き補償した後に、テロップに属する画素を抽出する方法を提案した。提案手法を用いてテレビ映像から、0.4pixel/sec 以下の誤差で移動速度を検出した。テロップ抽出結果を OCR を用いて認識したところ、54.6%の文字の認識に成功した。今後の課題は、他の速度検出手法との比較、細い線の抽出性能の向上である。

参考文献

- [1] 鷹尾誠一, 有木康雄, 緒方淳, “クロスメディア・パッセージ検索-テロップや CG フリップ文字列を検索質問とした発話文書に対する検索方式-,” 信学論, vol.J84-D-II, no.8, pp.1809-1816, Aug. 2001.
- [2] 堀修, 三田雄志, “テロップ認識のための映像からのロバストな文字部抽出法,” 信学論, vol.J84-D-II, no.8, pp.1800-1808, Aug. 2001.
- [3] L. Xu, and E. Oja, “Randomized Hough Transform(RHT): Basic mechanisms, algorithms and computational complexities,” CVGIP: Image Understanding, vol.57, no.2, pp.131-154, Mar. 1993.