社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

# ニュース映像中の同一シーン検出のための領域別照合手法の検討

熊谷はるか† 道満 恵介†† 出口 大輔††† 井手 一郎†† 村瀬 洋††

† 名古屋大学工学部 〒 464-8603 愛知県名古屋市千種区不老町
†† 名古屋大学大学院情報科学研究科 〒 464-8601 愛知県名古屋市千種区不老町
††† 名古屋大学情報連携統括本部 〒 464-8601 愛知県名古屋市千種区不老町
E-mail: †{kumagaih,kdoman}@murase.m.is.nagoya-u.ac.jp, ††ddeguchi@nagoya-u.ac.jp,
†††{ide,murase}@is.nagoya-u.ac.jp

**あらまし** 本稿では，同一のイベントを異なる視点で撮影した同一シーンを検出するための領域別照合手法の検討結果について報告する．近年では，大量に蓄積されたニュース映像を内容に基づいて扱う技術が必要となっている．本稿ではその中でも，同一シーン検出手法に注目する．本手法では，映像を人物領域と背景領域に分割し，それぞれの領域同士を独立に照合し，その結果を統合する．撮影視点の違いに対応するため，領域同士の照合では，それぞれの領域の性質に応じて，撮影視点の違いに頑健な特徴を用いる．人物領域の照合には，特徴点の不均一性パターンを用い，背景領域の照合には，色ヒストグラム特徴を用いる．実際に放送されたニュース映像で同一シーン検出実験を行い，提案手法の有効性を確認した．

**キーワード** 同一シーン検出，撮影視点の違い，領域別照合，ニュース映像

# A study on region-based matching for scene duplicate detection from news videos

Haruka KUMAGAI†, Keisuke DOMAN††, Daisuke DEGUCHI†††,

Ichiro IDE††, and Hiroshi MURASE††

† Nagoya University, Faculty of Engineering Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464–8603 Japan
†† Nagoya University, Graduate School of Information Science
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464–8601 Japan
††† Nagoya University, Information and Communications Headquarters
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464–8601 Japan
E-mail: †{kumagaih,kdoman}@murase.m.is.nagoya-u.ac.jp, ††ddeguchi@nagoya-u.ac.jp,
†††{ide,murase}@is.nagoya-u.ac.jp

**Abstract** In this paper, we report a study on region-based matching for detecting scene duplicates from news videos. Scene duplicates are videos taking the same event from different perspectives. Recently, technology to deal with the large amount of news videos based on their contents is needed. Among them, we are focusing on the scene duplicate detection problem. The proposed method first segments a video into the background and the person region, and then each region is matched independently, and finally the results are integrated. In order to deal with the different perspectives, the regions are matched with the features that are robust against the different perspectives according to the nature of each region. To match the person regions, the proposed method uses the pattern non-uniformity of the feature points. Meanwhile, to match the background region, the proposed method uses the color histogram feature. We conducted scene duplicate detection experiments using actual news videos. From the results, we confirmed the effectiveness of the proposed method.

**Key words** Scene duplicate detection, difference of perspective, region-based matching, news video

## 1. はじめに

近年，記憶装置や通信技術の発達により，放送映像を大量に蓄積できるようになった．放送映像には，スポーツ，ドラマ，バラエティ，アニメーションなど，さまざまな種類がある．その中でも，ニュース映像は実世界の出来事を記録しており，資料的観点から価値が高い．ニュース映像を資料として活用するためには，大量に蓄えたニュース映像を内容に基づいて扱う技術が必要となる．本研究では，そのような技術の中でも，映像の near-duplicate を検出する技術に注目する．Near-duplicate とは，同一のある時点・場所で起こった実世界の出来事（イベント）を撮影した映像，または，同じ対象を撮影した映像のことであり，ニュース映像の構造化 [1] などに利用できる．また near-duplicate は，以下の 3 種類に分けられる [2].

（1） Strict near-duplicate：同一のイベントを同じ撮影視点で撮影した映像のこと．図 1 に示すように，編集，字幕制作などによる違いがあり得るが，撮影視点は同じである．

（2） Object duplicate：同じ撮影対象について，異なるイベントを撮影した映像のこと．図 2 に示すように，同じ撮影対象を，同じ場所で，異なる時刻に撮影した映像のことであり，撮影視点の違いがあり得る．

（3） Scene duplicate：同じイベントを異なる撮影視点から撮影した映像のこと．図 3 に示すように，同じ場面を同じ時刻に，異なる視点から撮影した映像である．また，撮影時刻が同じであるが，異なるテレビ番組に利用されることから，放送される部分が異なることも多い．

本研究では，これらの中でも（1）と（3）の検出を目指す.（1）と（3）はある 1 つのイベントを撮影した映像である．ニュース映像では，ある 1 つのイベントに関するニュースでも，放送局ごとに使われる映像が異なったり，報道の仕方が異なる場合がある．このように，ニュース映像は放送局ごとに多様なため，（1）と（3） を検出できれば，ある 1 つのイベントを様々な視点のニュース映像から理解することが可能になる．

以降，2 で関連研究を紹介し，3 では提案手法について述べる．そして，4 で評価実験について述べ，その結果について考察する．最後に 5 で本報告をまとめる．

## 2. 関連研究

Near-duplicate の中でも，同一のイベントを同じ撮影視点で撮影した（1）を対象とする研究としては，長時間の映像から短時間の既知の映像を高速に検出することを目的とした柏野らの研究 [3] や，繰り返し複製されたり，加工されたりした映像を検出することを目的とした Law-To らの研究 [4] がある．

全ての near-duplicate を区別せずに検出する研究として，Ngo らの研究 [5] や，Zhang らの研究 [6] が挙げられる．これらの研究は，撮影視点の違いに頑健な特徴量を用いることによって，撮影視点の異なる（2）や（3）の検出を可能としている．

撮影時刻が異なる（2）と（3）を区別し，同一のイベントを撮影した（1）と（3）を検出することを目的とした研究として，瀧本らの研究 [7] や武らの研究 [2] がある．瀧本らは，映像中のフラッシュ光を利用した検出手法を提案している．しかし，この手法は，フラッシュの発生に依存するため，フラッシュが映っていない映像は検出できない．武らは，映像中の人物の特徴点の不均一性パターンに基づいて（1）と（3）を検出する手法を提案している．この特徴点の不均一性パターンは，人物の動きの緩急を表す特徴である．人物の動きの緩急は，撮影視点の違いに頑健であり，また，同じ撮影対象でも撮影時刻によって変化する．したがって，撮影時刻の異なる（2）と（3）を区別して検出することができる．しかしこの手法は，図 4 に示すような，背景に特徴点が集まり，人物に特徴点が少なくなる映像は検出できない．背景に特徴点が集まるのは，背景に文字や絵があるような場合である，このようなニュース映像は，記者会見映像などでたびたび見られるので，実際のニュース映像からの検出にはこの問題の解決が必要である．

図 1　Strict near-duplicate の例

図 2　Object duplicate の例

図 3　Scene duplicate の例

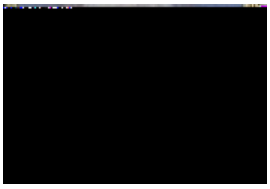

図 4　人物の特徴点が少ない映像 [2]

## 3. 提案手法

前節で述べた内容をふまえ，本稿では，strict near-duplicate と scene duplicate を合わせて同一シーンと呼び，ニュース映像から同一シーンを検出する手法を提案する．

### 3.1 手法の概要

提案手法の流れを図 5 に示す．まず，入力された 2 つのショットを，それぞれ人物領域と背景領域に分割する．そして，領域

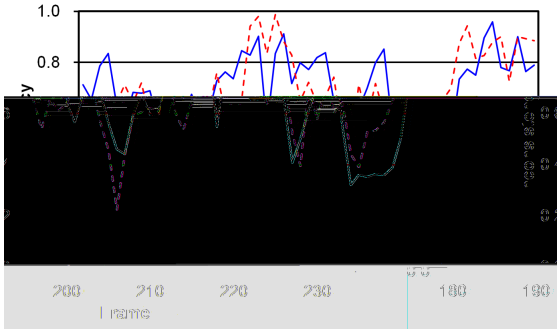


図 7　Inconsistency の変化の例

行列 M の固有値を $\lambda_1$，$\lambda_2$，$\lambda_3$（$\lambda_1 \geqq \lambda_2 \geqq \lambda_3$）とし，行列 $M^{\diamond}$ の固有値を $\lambda_1^{\diamond}$，$\lambda_2^{\diamond}$（$\lambda_1^{\diamond} \geqq \lambda_2^{\diamond}$）とすると，inconsistency は，式 (3) を用いて $[0, 1]$ の間の連続的な実数として求めることができる．

$$\text{Inconsistency} = \frac{\lambda_2 \cdot \lambda_3}{\lambda_1^{\diamond} \cdot \lambda_2^{\diamond}} \tag{3}$$

実際のショットにおける，ある特徴点に関する inconsistency の変化を図 7 に示す．Inconsistency は，特徴点の動きの緩急を表す．値が大きいフレームは，特徴点の動きの向きや速度が大きく変化するフレームを表し，逆に値が小さいフレームは，動きの向きや速度が一定であるフレームを表す．

次に，求めた inconsistency を基に，discontinuity を計算する．Discontinuity とは，ショット間の照合に用いるもので，discontinuity が 1 となるフレームに関して，ショット間の inconsistency の類似度を計算する．Discontinuity は，inconsistency が時間方向のある幅のウインドウ内の値と比較した際に極大値をとるフレームにおいて 1 になる．つまり，あるフレームにおいて，ウインドウ内で inconsistency が極大値をとるならば，discontinuity を 1 とし，そうでなければ 0 とする．

最後に，ショット間の類似度を求める．1 つのショットは複数の特徴点軌跡によって表されるため，まず，軌跡間の類似度を求める．軌跡間の類似度は，discontinuity が 1 になるフレームの前後 $w$ フレームにおいて，両軌跡の inconsistency の相関係数を計算して求める．この $w$ フレームは，discontinuity を求める際のウインドウとは独立である．そして，計算された全ての相関係数の平均値を軌跡間の類似度として定義する．$i$ 番目のショット $S_i$ が $n_i$ 本の軌跡 $T_i^j$（$j = 1, \cdots, n_i$）を持つとすると，2 つの軌跡 $T_1^j$ と $T_2^k$ の間の類似度 $\mathrm{Sim}(T_1^j||T_2^k)$ は式 (4)，(5) のように計算される．なお説明の簡略化のため，ここでは軌跡のインデックスを省略する．

$$\mathrm{Sim}(T_1||T_2;\tau) = \frac{\sum_t D \times \mathrm{NCC}\big(c(t;T_1), c(t-\tau;T_2); t-w, t+w\big)}{\sum_t d(t;T_1) + d(t-\tau;t_2)} \tag{4}$$

$$\mathrm{NCC}\big(c(t;T_1), c(t;T_2); t_1, t_2\big) = \frac{\sum_{t=t_1}^{t_2} \big(c(t;T_1) - \overline{c(T_1)}\big)\big(c(t;T_2) - \overline{c(T_2)}\big)}{\sqrt{\sum \big(c(t;T_1) - \overline{c(T_1)}\big)^2 \sum \big(c(t;T_2) - \overline{c(T_2)}\big)^2}} \tag{5}$$

図 8　映像の明るさが異なる同一シーン例

$\alpha$，$\beta$ をそれぞれショットもしくは軌跡とすると，$\mathrm{Sim}(\alpha||\beta)$ は $\alpha$ と $\beta$ の類似度を表す．$\tau$ は時間的なオフセットを示す．同一シーンは，放送される部分が各番組によって異なる場合がある．よって，そのような時刻のずれを吸収するために，オフセットごとに類似度を計算する．ここで，$d(t;T)$，$c(t;T)$ は各々軌跡 $T$ の $t$ 番目のフレームにおける discontinuity の値と軌跡 $T$ の $t$ 番目のフレームにおける inconsistency の値を示す．$\mathrm{NCC}(c(t;T_1), c(t;T_2); t_1, t_2)$ は，$t_1 \leqq t \leqq t_2$ において，inconsistency $c(t;T_1)$，$c(t;T_2)$ の間の正規化された相互相関係数である．また，$\overline{c(T_i)}$ は，$c(t;T_i)$ の平均である．

以降，便宜上，時間的なオフセット $\tau$ を省略する．式 (6) に示すように，片方のショット $S_1$ 中の $j$ 番目の軌跡 $T_1^j$ と，もう一方のショット $S_2$ との類似度 $\mathrm{Sim}(T_1^j||S_2)$ には，軌跡 $T_1^j$ に対して類似度が最も高い軌跡 $T_2^k$ の類似度を用いる．さらに，式 (7) に示すように，ショット $S_1$ に対するショット $S_2$ の類似度 $\mathrm{Sim}(S_1||S_2)$ には，式 (6) で計算された軌跡とショットの類似度の高い方から $\rho\%$ の平均値を用いる．最後に，2 つのショット $S_1$ と $S_2$ の間の類似度 $R_{\mathrm{face}}(S_1, S_2)$ には，式 (8) に示すように，$S_1$ 対する $S_2$ の類似度と $S_2$ 対する $S_1$ の類似度の平均値を用いる．

$$\mathrm{Sim}(T_1^j||S_2) = \max_k \mathrm{Sim}(T_1^j||T_2^k) \tag{6}$$

$$\mathrm{Sim}(S_1||S_2) = \operatorname*{avg}_{\mathrm{top}\rho\%} \mathrm{Sim}(T_1^j||S_2) \tag{7}$$

$$R_{\mathrm{face}}(S_1, S_2) = \frac{1}{2}\big(\mathrm{Sim}(S_1||S_2) + \mathrm{Sim}(S_2||S_1)\big) \tag{8}$$

### 3.3.2　背景領域の照合

本研究では，背景領域の類似度として，色ヒストグラム間の距離を用いる．ここで図 8 に示すように，ニュース映像は，録画状況や各放送局の規定によって同一シーンでも映像の明るさが異なる場合がある．この明るさの違いによる検出漏れや誤検出を避けるために，色ヒストグラムには HSV 値の色相（H）と彩度（S）のみを用いる．まず，領域分割されたショットの最初のフレームから色相と彩度を求め，2 次元ヒストグラムを計算する．次に，式 (9) に示すように，ヒストグラム間の Bhattacharyya 距離を求め，1 から引くことで背景領域の類似度 $R_{\mathrm{bg}}$ を計算する．

$$R_{\mathrm{bg}} = 1 - \sqrt{1 - \sum_{u=1}^{m} \sqrt{p^{(u)} q^{(u)}}} \tag{9}$$

ここで，$p^{(u)}$，$q^{(u)}$ はそれぞれ正規化ヒストグラムにおける $u$ 番目のビンの値を表し，$m$ はヒストグラムのビン数である．

## 3.4　個別照合結果の統合

人物領域の類似度 $R_{\mathrm{face}}$ と背景領域の類似度 $R_{\mathrm{bg}}$ から，入力

表 1　実験データとして使用した放送局と番組，放送時刻

| 放送局 | 番組名 | 放送時刻 |
|---|---|---|
| NHK 総合 | ニュース | 12:00 〜 12:20 |
| 東海テレビ | FNN スピーク | 11:30 〜 11:55 |
| 中京テレビ | NNN ストレイトニュース | 11:30 〜 11:55 |
| CBC テレビ | JNN ニュース | 11:30 〜 11:55 |
| メ〜テレ | ANN ニュース | 11:45 〜 12:00 |

ショット $S_i$，$S_j$ 間の類似度 $R_{\mathrm{fusion}}$ を次式で計算する．

$$R_{\mathrm{fusion}}(S_i,S_j)=\alpha R_{\mathrm{face}}(S_i,S_j)+(1-\alpha)R_{\mathrm{bg}}(S_i,S_j) \quad (10)$$

ここで，$\alpha$ は 0 から 1 の値をとる重み係数を表す．

入力ショットが複数の場合，全てのショット間で類似度を計算する．そして，ショット間の類似度が最も大きいものを選択する．また，入力ショットに同一シーンが存在しない可能性があるため，類似度がしきい値以上のものを同一シーンと判別する．しきい値は，線形判別分析法によって決定する．

## 4. 実験・考察

実際に放送されたニュース映像に本手法を適用した実験の結果について述べる．

### 4.1 実験用データ

本実験では，入力ニュース映像として，2011 年 12 月 9 日に放送されたニュース映像を使用した．使用した放送局と番組，放送時刻を表 1 に示す．また，入力ニュース映像のフレームレートは 29.97 [frames/sec]，解像度は 1,440 × 1,080 [pixels] であった．これらの映像から，人物を中心に撮影しているショットを人手で計 38 ショット切り出した．この 38 ショットのうち，同一シーンが存在するショットは 24 ショット，存在しないショットは 14 ショットであった．なお，人物領域は，同一シーン検出の検出精度への影響を排除するため，全て人手で切り出した．

### 4.2 実験方法

提案手法の有効性を評価するため，以下の手法による同一シーン検出の検出精度を比較した．

**提案手法：** 入力ショットを人物領域と背景領域に分割し，それぞれを個別に照合し，その結果を統合した．人物領域の照合では特徴点の不均一性パターン [2] を用い，背景領域の照合では色ヒストグラム特徴を用いた．

**比較手法 1：** 入力ショットを人物領域と背景領域に分割せず，提案手法の人物領域の照合で用いる特徴を用いて照合した．また，入力映像の解像度を 360 × 270 [pixels] とした．これは，KLT トラッカによる特徴点抽出は，高解像度画像を入力した場合に長時間かかってしまうためである．

**比較手法 2：** 入力ショットを人物領域と背景領域に分割せず，色ヒストグラム特徴を用いて照合した．色ヒストグラムの作り方と，色ヒストグラム間の距離の計算方法は，提案手法の背景領域の照合で用いる方法と同じとした，

**比較手法 3：** 入力ショットを人物領域と背景領域に分割し，人物領域のみを照合した．照合方法は，提案手法の人物領域の照合方法と同じとした．

**比較手法 4：** 入力ショットを人物領域と背景領域に分割し，背景領域のみを照合した．照合方法は，提案手法の背景領域の照合方法と同じとした．

表 2　実験結果

| 手法 | 領域 | 特徴 | 再現率 | 適合率 | F 値 |
|---|---|---|---|---|---|
| 提案手法 | 人物領域＋背景領域 | 動き＋色 | 0.79 | 0.76 | 0.78 |
| 比較手法 1 | 全体 | 動き | 0.16 | 0.15 | 0.16 |
| 比較手法 2 | 全体 | 色 | 0.75 | 0.75 | 0.75 |
| 比較手法 3 | 人物領域 | 動き | 0.33 | 0.34 | 0.34 |
| 比較手法 4 | 背景領域 | 色 | 0.71 | 0.74 | 0.73 |

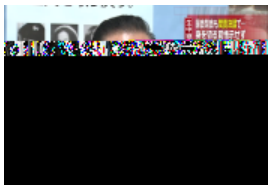

図 9　検出成功例

図 10　未検出例

図 11　誤検出例

上記の 5 種類の手法を 2 分割交叉検定により評価した．具体的には，実験用データセットを 2 つに分け，各セットにつき同一シーンが存在するショットが 12 ショット，存在しないショットが 7 ショットとなるように手作業で分けた．評価基準としては，適合率，再現率，F 値を調べた．

### 4.3 実験結果

各手法による実験結果の適合率と再現率，F 値の平均値を表 2 に示す．また，図 9 に検出成功例を，図 10 に未検出例を，図 11 に誤検出例を示す．全ての手法を比較すると，提案手法が適合率，再現率，F 値の平均値に関して最も高くなり，提案手法の有効性を確認できた．

### 4.4 考察

未検出は，図 10 に示したショットのみであった．これは，背景領域に，人物の背広の部分が含まれたため，その割合の違いによって色ヒストグラムが変化し，類似度が小さくなったためだと考えられる．また誤検出は，全て図 11 に示したような同じ場所だが異なる時刻に撮影したショットであった．これは，背景領域の類似度が高くなり，人物領域の類似度の影響が小さくなったために起きたと考えられる．

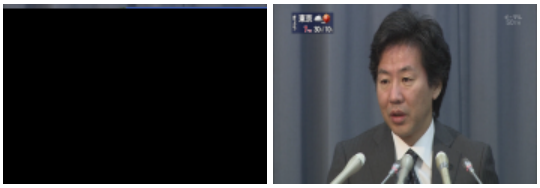

図 12　同じ撮影対象だが撮影時刻の異なるショット

また，比較手法 1 と比較手法 3 から，人物の動きの緩急を表す特徴量である特徴点の不均一性パターンは，人物領域のみに適用するとより効果的であることが確認できた．比較手法 1 と比較手法 3，つまり特徴点の不均一性パターンのみを用いた手法は，表 2 に示すように，ほかの手法と比べ，F 値の平均値が低かった．これは，実験に用いたショットの長さがさまざまであったために，多くのショットが短いショットに対応付けられたためである．人物領域の類似度の計算は，ショットの時間が短い方に合わせて計算され，一方のショットの時間が短いほど，照合される範囲が短くなる．つまり，同一シーンではない時間の長いショット同士は，人物の動きの緩急が異なる部分が時間の長さの分だけ多くなるが，長いショットと短いショットでは，異なる部分が少なくなり，そのために類似度が大きくなり，対応付けられやすくなる．このような理由で，誤対応が多くなったのだと考えられる．ただし，特徴点の不均一性パターンを用いる手法は，同じ撮影対象で異なるイベントを撮影した映像を見分けることが目的であり，本実験でも，図 12 のような同じ撮影対象だが撮影時刻の異なるショットを見分けることはできた．このような特性を活かすためには，背景領域の照合結果を人物領域の照合結果を用いて絞り込むなど，人物領域と背景領域の照合結果の統合方法の更なる工夫が必要である．

色ヒストグラムのみを用いる比較手法 2 と比較手法 4 を比較すると，入力ショットを人物領域と背景領域に分割せずに照合する比較手法 2 の方が，背景領域のみを用いて照合する比較手法 4 よりも，F 値の平均値が大きかった．本実験では．人物領域を矩形で分割しているために，人物領域にも背景が含まれる．そのため，この結果が，色ヒストグラムによる照合に重要な情報が，人物領域の人物部分に含まれているからなのか，人物領域の背景部分に含まれているからなのか，領域の分割方法を変化させるなどして調査する必要がある．

## 5. む す び

本稿では，ニュース映像における人物領域と背景領域の個別照合による同一シーン検出手法を提案した．また，提案手法の有効性を確認するために，実際に放送されたニュース映像からの同一シーン検出を行った．

提案手法では，撮影視点の違いを解決するために，映像を人物領域と背景領域に分割し，それぞれの領域を撮影視点の違いに頑健な特徴を用いて照合し，その結果を統合して同一シーンを検出する．人物領域と背景領域に分割することにより，それぞれの領域が互いに影響されず，独立して照合されるように工夫した．人物領域の照合には，人物の動きの緩急を表す特徴点の不均一性パターンを用いた．人物の動きは撮影視点の違いに影響を受けにくいため，これにより撮影視点の違いに頑健な対応付けが可能となる．また，人物の動きの緩急のみでは，動きの緩急は似ているが，映像的な見た目が大きく異なる映像は区別できない．そこで，背景領域には，色ヒストグラム特徴を用いた．この特徴は，撮影視点の違いに頑健であり，映像的な見た目が大きく異なる映像を区別することができる．

提案手法の有効性を確認するために，実際のニュース映像を用いた同一シーン検出実験を行った．比較手法は，映像の領域分割を行う場合と行わない場合，特徴点の不均一性パターンのみを用いる場合と色ヒストグラム特徴のみを用いる場合を組み合わせた 4 通りの手法とした．その結果，提案手法の精度が最も高く，提案手法の有効性が確認できた．今後は，最適な個別照合結果の統合方法や領域分割方法の検討をしていく．

**謝辞**　本研究の一部は科学研究費補助金及び国立情報学研究所との共同研究による．

## 文　献

[1] X. Wu, C.-W. Ngo and A.G. Hauptmann, “Multimodal news story clustering with pairwise visual near-duplicate constraint,” IEEE Transactions on Multimedia, vol. 10, no. 2, pp. 188–199, Feb. 2008.

[2] 武 小萌, 瀧本 政雄, 佐藤 真一, 安達 淳, “特徴点軌跡の不均一性パターンに基づいた同一場面映像検出,” 電子情報通信学会論文誌（D）, vol. J92-D, no. 8, pp. 1135–1165, Aug. 2009.

[3] 柏野 邦夫, 黒住 隆行, 村瀬 洋, “ヒストグラム特徴を用いた音や映像の高速 AND/OR 探索,” 電子情報通信学会論文誌（D-II）, vol. J83-D-II, no. 12, pp. 2735–2744, Dec. 2000.

[4] J. Law-To, O. Buisson, V. Gouet-Brunet and N. Boujemaa, “Robust voting algorithm based on labels of behavior for video copy detection, ”Proc. 14th ACM International Conference on Multimedia, pp. 835–844, Oct. 2006.

[5] C.-W. Ngo, W. Zhao and Y.-G. Jiang, “Fast tracking of near-duplicate keyframes in broadcast domain with transitivity propagation,” Proc. 14th ACM International Conference on Multimedia, pp. 845–854, Oct. 2006.

[6] D.-Q. Zhang and S.-F. Chang, “Detecting image near-duplicate by stochastic attributed relational graph matching with learning,” Proc. 12th ACM International Conference on Multimedia, pp. 877–884, Oct. 2004.

[7] 瀧本 政雄, 佐藤 真一, 坂内 正夫, “大容量放送映像アーカイブからの同一フラッシュシーン映像の発見,” 電子情報通信学会論文誌（D）, vol. J89-D, no. 12, pp. 2699–2709, Dec. 2006.

[8] R. Lienhart and J. Maydt, “An extended set of Haar-like features for rapid object detection,” Proc. 2002 IEEE International Conference on Image Processing, vol.1, pp.900–903, Sep. 2002.

[9] A. Kuranov, R. Lienhart and V. Pisarevsky, “An empirical analysis of boosting algorithms for rapid objects with an extended set of Haar-like features,” Intel Tech. Rep., MRL-TR-July02-01, Jul. 2002.

[10] C. Tomasi and T. Kanade, “Detection and tracking of point features,” Carnegie Mellon University Tech. Rep., CMU-CS-91-132, Apr. 1991.

[11] E. Shechtman and M. Irani, “Space-time behavior based correlation,” Proc. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.405–412, Jun. 2005.