

# 画像同定技術：Visual Signature

工藤大樹\*  
西川博文\*

Image Identification Technique: Visual Signature

Daiki Kudo, Hirofumi Nishikawa

## 要 旨

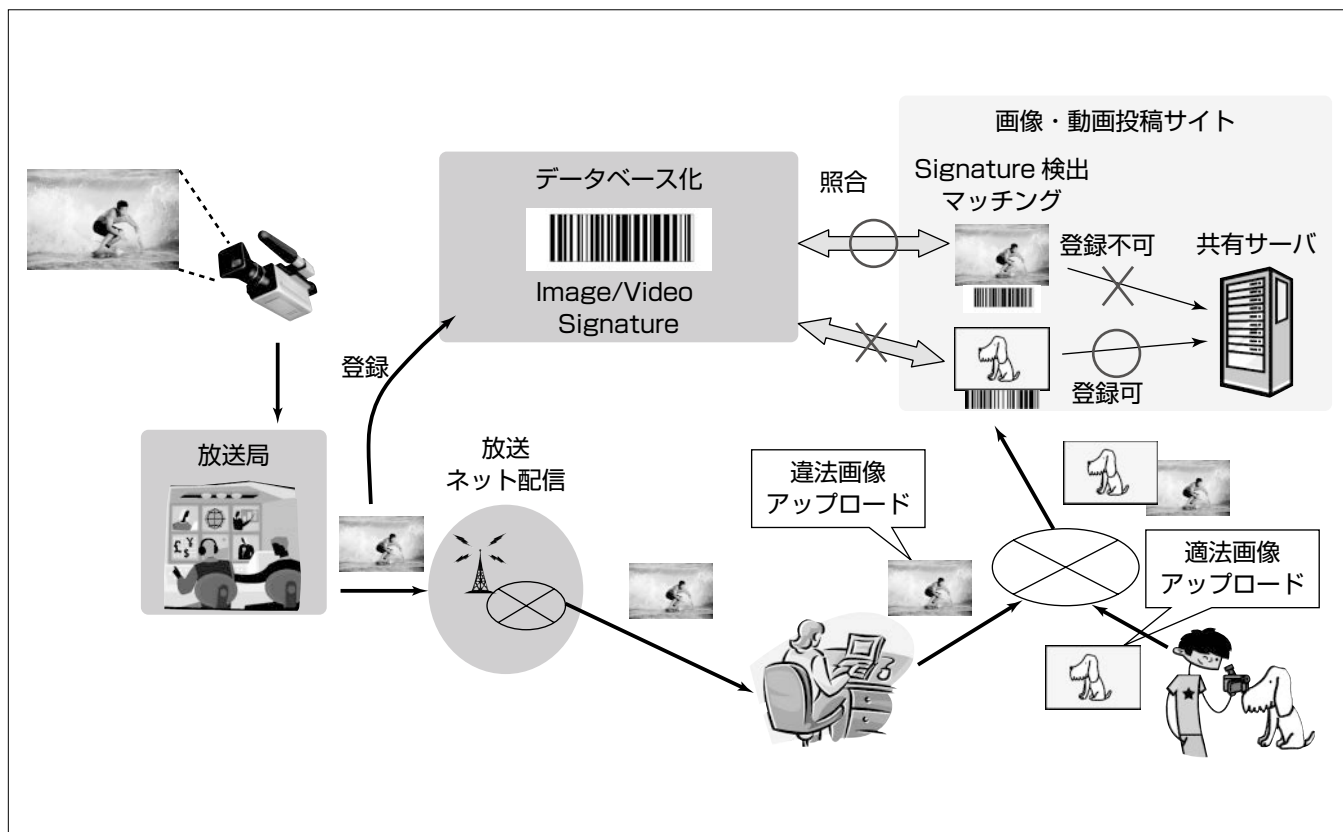
インターネットは、回線の高速化に加え、デジタルカメラ、パソコン、携帯電話の普及によって、画像映像の膨大な流通インフラとなっている。さらに、スマートフォンの普及によって、画像の発信・活用がだれにでも可能な世界が実現されていくだろう。しかし、所望の画像映像コンテンツの効率的な検索は困難であり、有効な検索・管理手法への要求は、さらに高まっている。

そこで、タイトルなどのテキスト情報に加え、画像そのものの特徴量から、画像コンテンツの検索管理を行う方法が検討されてきている。ISO (International Organization for Standardization) / MPEG (Moving Picture Experts Group) では、この画像特徴量の記述方法の標準化を行い、2002年にMPEG-7標準 (ISO/IEC 15938シリーズ) を発行し

た。このMPEG-7の拡張として画像の同一性の検出を目的に定義されたのが、三菱電機が標準化に大きく寄与したVisual Signatureである。

Visual Signatureは、画像の空間的な特徴を効率的に表現するもので、回転・拡大・縮小等の画像編集後も同一性の判断が可能である。部分切り出しへの耐性を持つ静止画向けのImage Signatureと、時間軸上での編集への耐性を持つ動画向けのVideo Signatureがある。

この技術によって、違法画像コンテンツの検出など、画像コンテンツ管理のほか、実世界の画像からインターネット上の同一画像を検索する画像検索型のネットショッピングや、ユーザーに必要な情報を提供するサービス等、幅広い画像利用サービスが実現できる。



## Visual Signatureによる違法コンテンツフィルタリング

あらかじめ、著作権のある画像に対してVisual Signature 特徴量を抽出し、データベース化しておく。ユーザーが画像をアップロードした場合、まずマッチングサーバに送られ、データベース内の画像とマッチングが行われる。その結果、マッチングしなかった、すなわち、著作権ありとして登録されていなかった画像のみ蓄積サーバに登録される。

Visual Signature の高速マッチングによって、実用的な時間で処理可能となる。

## 1. ま え が き

現在、インターネットには多くの画像コンテンツがあふれている。回線の高速化に加え、デジタルカメラ、パソコン、携帯電話等、画像が簡単に扱える機器の普及によって、コンテンツ共有サービスは拡大し、インターネットはマルチメディアコンテンツの流通インフラとなっている。今後、クラウドベースのストレージサービスも普及し、個人のコンテンツもSNS(Social Networking Service)などを通じて共有することが当然となるだろう。スマートフォンやタブレット端末にも、十分なローカルストレージと無線ネットワーク環境が備わり、大量のコンテンツに対して“いつでも”“どこでも”アクセス可能な環境が整ってきている。環境がサービスを普及させ、サービスの普及がさらにサービスを魅力的なものとし、インターネット上のマルチメディアコンテンツは膨大なものになると考えられる。

このような環境では、ユーザーが所望のコンテンツを素早く簡単に見つけること、また、権利侵害など違法利用を防止することは極めて重要であり、これを実現するための画像検索・管理技術への期待はますます高まっている。

画像情報管理では、作成者などのテキストデータに加え、画像そのものの特徴をメタデータとすることが検討されてきた。MPEG-7(ISO/IEC 15938シリーズ)は、MPEG-2/4などの画像圧縮技術と異なり、マルチメディアコンテンツの検索を行うためのメタデータの国際標準であり、検索に用いる特徴量及び記述子の標準化が行われてきた。

近年、このMPEG-7で画像の同一性の判定を行うVisual Signature(画像同定)技術の標準化が行われた。この標準化で、当社は、その提案技術が標準方式に採用されるなど、中心的な役割を果たしてきた<sup>(1)</sup>。本稿では、このVisual Signature技術について述べ、利用分野を展望する。

## 2. Visual Signature

### 2.1 画 像 同 定

画像共有サイトには、権利者の許可なく画像が投稿、共有される場合がある。これらの画像を容易に検出するのが画像同定技術である。

一般に、デジタルデータの同一性を判断する方法として、SHA(Secure Hash Algorithm)に代表されるハッシュ関数がある。ハッシュ関数とは、デジタルデータのビット列を要約する関数である。内容が1ビットでも異なると、全く異なる値になるため、効率的にデジタルデータの同一性を確認できる。画像データの場合、図1上部に示すように、回転やモノクロ化などの処理を施しても、人間の目には同一画像として認識される。しかし、これらの画像はすべてビット列が異なるため、ハッシュ値が異なってしまう。このように、画像同定では、編集・加工に対する頑健性が要

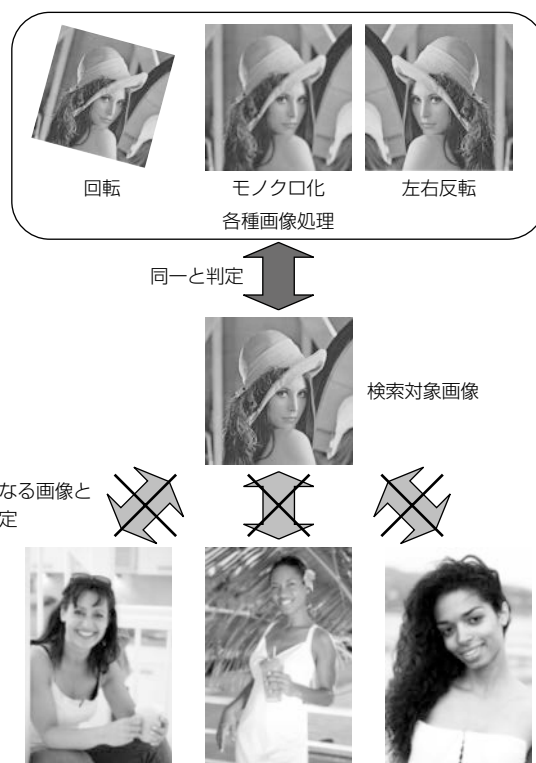


図1. 同一の画像

求され、ハッシュ関数では対応できない。

Visual Signatureは、これらの処理を行っても同一の画像と判断できることを目的として開発された。“女性の画像”を検索するための類似画像検索技術とは異なり、Visual Signatureでは図1下部の画像は別画像と判定されなくてはならない。

### 2.2 Image SignatureとVideo Signature

Visual Signatureには、静止画向けのImage Signature<sup>(2)</sup>と、動画像向けのVideo Signature<sup>(3)</sup>の2種類が存在する。

Image Signatureはトレース変換<sup>(4)</sup>を基にした特徴量であり、2009年に標準化された。輝度情報を空間的周波数で表現しており、特徴量間の独立性が極めて高く、ほかの画像を誤って同一と判断するのは100万ペア中で0.05ペアである。また、画像特徴を空間的な分布で表現するため、画像のクロッピング(一部の切り取り)など、大幅な改変処理が施された後でも94%以上を同一と判定する高い頑健性を持つ<sup>(1)</sup>。

一方、動画像データに対する同定を目的としたVideo Signatureは、2010年に標準化された。Image Signature同様、高い独立性・頑健性を持つ。

## 3. Image Signature

静止画の同定では、画像全体の特徴を高速に比較するGlobal Signatureと、特徴点を抽出し比較するLocal Signatureの2種類のマッチング手法が用意されており、用途に応じて利用可能である。

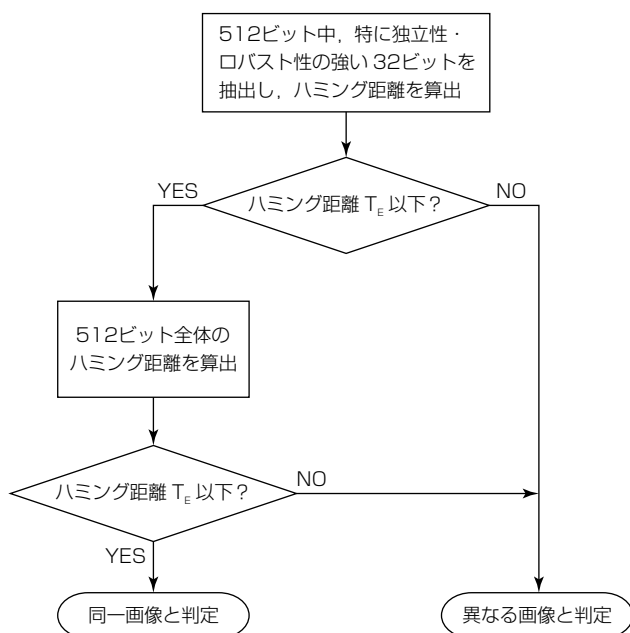


図 2. Global Signature 高速マッチングのフローチャート

### 3.1 Global Signature 高速マッチング

Global Signature は画像全体を 512 ビットの特徴量で表現する。通常、2 特徴量間のハミング距離 (2 つのビット列中、対応する位置にある異なったビットの個数) を特徴量距離とし、高速なマッチングが可能である。

しかし、比較対象画像数が膨大な場合など、より優れたユーザーインターフェースを実現するために、さらに高速なマッチング方法が考案されている。

その手順を次に示す (図 2)。

① 512 ビットの特徴量から 32 ビットを選択する。

② 選択した 32 ビット間のハミング距離を算出、閾値 (しきいち) 判定を行う。閾値よりハミング距離が小さい場合のみ次の③に進む。

③ 従来の 512 ビットのハミング距離を算出し、閾値判定処理を行う。

すなわち、部分特徴量で予備判断をするマッチング手法である。利用する 32 ビットは、独立性・頑健性の高いものが標準化の過程で実験的に選択、規格化されている<sup>(5)</sup>。

この高速化手法によって、Global Signature マッチング速度は 7,000 万枚/秒 (Intel Core2Duo<sup>(注1)</sup> 3 GHz 使用時) に達し、マッチング精度の低下は 1 % 程度であり、実用上問題ないレベルである。

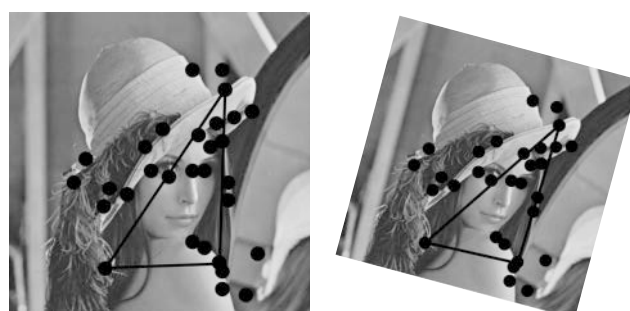
(注 1) Core2Duo は、Intel Corp. の登録商標である。

### 3.2 Local Signature マッチング

Local Signature は、画像の特徴点ごとに局所的なトレース変換を行い、特徴量とするものである。図 3 に特徴点の例を示す。このマッチングでは、特徴量距離の近い特徴点が幾何学的に類似しているかどうかを次の手順で判定する。



図 3. Local Signature の特徴点



三角形の辺の比を比較

図 4. 3 特徴点ペアによる幾何学的制限

① 特徴点ペアを 3 つ選択する

特徴量距離が閾値を下回るペアを 3 つ選択する。

② 3 特徴点ペアによる幾何学的制限 (図 4)

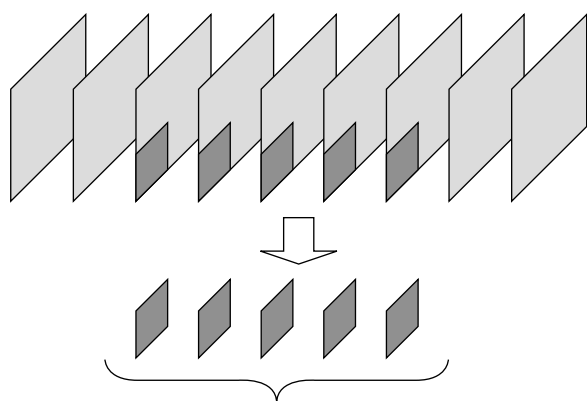
3 つの特徴点ペアから三角形を作成する。それぞれの画像中の 2 つの三角形の辺の大きさ ( $a_1, b_1, c_1$ ) と ( $a_2, b_2, c_2$ ) の比を比較し、類似の場合のみ次の③に進む。

③ 4 特徴点ペアによる幾何学的制限

特徴量距離の近い特徴点ペアをもう一つ選ぶ。4 ペアから 4 つの三角形が得られる。対応する各三角形の面積比を基に最終的なマッチング判定を行う。

この Local Signature の採用によって、画像のクロッピングに対する頑健性が強化されている。クロッピング後、画像の幅及び高さが 70% 以上保持されている場合、94% のマッチング精度を実現している。

ハミング距離を利用する Global Signature に比べてマッチング速度は劣るものの、10 万枚/秒と十分に高速である。



Picture-in-Pictureの映像のみから抽出

図 5. Picture-in-Pictureからの抽出

#### 4. Video Signature

Image Signature同様、各種改変処理が行われた映像に対しても、同一性を保持する特徴量である。図 5 に示すように、映像中に別の映像を埋め込んだPicture-in-Pictureの子画面部分の検索に対応するため、映像中の特定領域及び特定時間からの抽出も可能である。

Video Signatureは次の 3 つの要素で構成する。

##### (1) Frame Signature (画像一枚の特徴量)

Frame SignatureはImage Signatureと同様に画像一枚の特徴量である。映像中の全フレームに付与されることから、Image Signatureに比べて小さな演算量で抽出できる。まず、図 6 に示すように画像を様々な矩形(くけい)領域に分割する。次に、領域ごとの輝度平均値や、領域間の輝度差分値から380次元のベクトルを得る。取得したベクトルは可変長符号化され、最終的に608ビットの特徴量となる。

##### (2) Word (Frame Signatureのサブセット)

WordはFrame Signatureのサブセットであり、380次元のベクトルから25個の要素を計40ビットで表現する。

##### (3) Bag-Of-Words (時間方向の特徴を代表する要素)

Bag-Of-WordsはWordの代表的要素を時間方向にまとめた特徴量である。図 7 に示すように、90フレームを一つの抽出単位とし、さらに45フレームずつ互い違いに抽出を行う。映像の一部をカットする編集に対してもマッチングが可能である。

Video Signatureでは、はじめにBag-Of-Wordsによって時間方向で一致する位置を見つけ、その後、Frame Signatureによって画像一枚ごとに照合を行うことでマッチングを実施する。

#### 5. む す び

画像が含まれるマルチメディアコンテンツに対する検索技術への要求は高いが、技術的には困難な課題があった。しかし、大量のコンテンツに対して“いつでも”“どこでも”

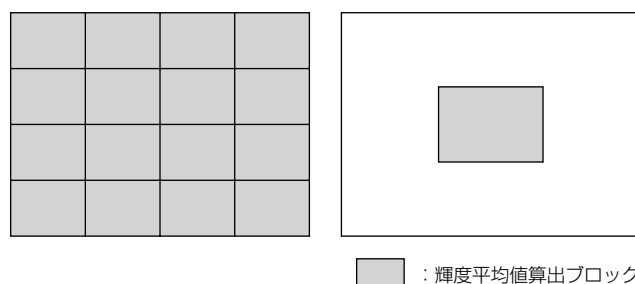


図 6. 領域分割例

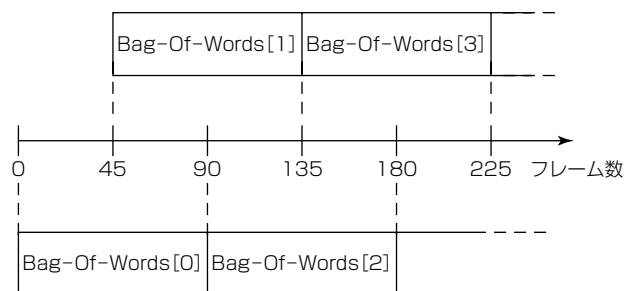


図 7. Bag-Of-Wordsの抽出

アクセス可能な環境が整い、様々なコンテンツ提供サービスが急速に拡大していく中、画像映像に対する検索・管理の重要性はますます高まっている。

その中で、画像同定方法として当社が国際標準化に大きく寄与したVisual Signatureは、画像の加工編集に対して頑健な特徴を少ない情報量で表現することができ、高精度・高速マッチングを実現している。

また、現在MPEG-7ではCDVS (Compact Descriptors for Visual Search) と呼ばれるモバイル端末向けの画像検索特徴量の標準化が始まっている。

この技術によって、違法コンテンツの検出に留まらず、画像情報からインターネット上の同一画像を検出することで、画像検索型のショッピング、情報検索、さらには、スマートフォンで撮影した看板やロゴなどに関連する情報を提供する情報サービス等、様々なアプリケーションへの適用が可能である。

#### 参 考 文 献

- (1) 西川博文, ほか: MPEG-7技術-Image Signature: 画像同定技術-, 三菱電機技報, **82**, No. 12, 763~766 (2008)
- (2) ISO/IEC 15938-3:2002/Amd 3:2009
- (3) ISO/IEC 15938-3:2002/Amd 4:2010
- (4) Kadyrov, A., et al.: The Trace Transform and Its Applications, IEEE Trans. Pattern Analysis and Machine Intelligence, **23**, No. 8, 811~828 (2001)
- (5) ISO/IEC TR 15938-8:2002/Amd 5:2010