# 映像メタデータ自動付与実現のための映像ショット・静止画像マッチング手法の一検討

関野 真洋† 青木 輝勝‡ 沼澤 潤二‡

東北大学大学院情報科学研究科† 東北大学電気通信研究所‡

## 1. はじめに

近年，データの圧縮技術やネットワーク関連技術，情報ストレージ技術の進展を背景に，映像コンテンツが一般的に広く利用されてきている．しかし，ユーザが映像コンテンツの取得，蓄積を容易に行うことができる一方，映像データが大量になることで，ユーザが閲覧したいコンテンツにたどり着くことが困難になっている．膨大な映像コンテンツの中から希望のコンテンツを高速検索するには，映像コンテンツへのメタデータ付与が必要であるが，メタデータ付映像コンテンツはごく一部に限られている．一方，Web 上のほとんどの静止画像には，検索エンジンにて自動的にメタデータが付与されている．従って，映像コンテンツと Web 上の静止画像との対応付けができれば映像コンテンツへのメタデータ自動付与が可能となる（図 1）．これを実現するため，連続して撮影された映像ショットと静止画像間の対応付けについて検討する．

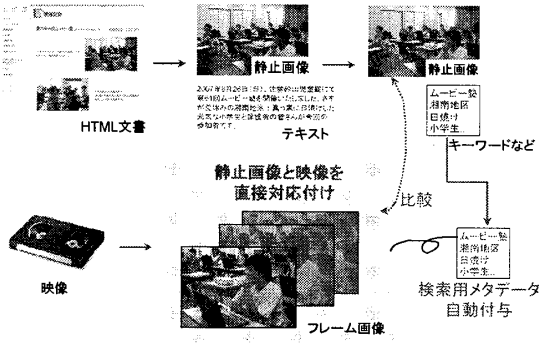


図 1 システム概念図

## 2. 関連研究

Web 上静止画像に付与されているメタデータは，HTML 文書内のキーワードを元に付与される．文書内にはその静止画像における撮影対象に関連した記述がなされていると考えられる．したがって，静止画像と類似した部分領域をもつフレーム画像を含むショットとの対応付けを行うことができれば，ショットに対して撮影対象に関連したキーワードを付与できると考えられる．

A Study on a New Image – Video shot Matching Method for Automatic Metadata Generation of Video Content

† Masahiro Sekino, Graduate School of Information Sciences, Tohoku University

‡ Terumasa Aoki, Junji Numazawa, Research Institute of Electrical Communication, Tohoku University

映像内に含まれる特定のオブジェクト領域の有無を検出する技術として，画像内の特徴点についての画像特徴量の一種である SIFT 特徴量[1]を用いた看板検出を行ったものがある[2]．この研究では，検出したいオブジェクトのみが存在する画像に含まれる SIFT 特徴量が，検出対象となる映像のフレーム画像に含まれているかを判定することで看板検出を行っている．SIFT 特徴量は局所特徴量であり，特徴量が画像間で一致することは，両画像において局所領域の一致があったことを意味する．一方の画像は看板のみが含まれる画像であるため，特徴量が一致する特徴点が多ければ多いほど，フレーム画像内に看板に類似した領域が多いといえる．

## 3. 提案手法

本システムにおいては，Web 上静止画像，映像ショットとのマッチングを行うため，映像ショットに含まれる複数のフレーム画像全体から抽出できる全 SIFT 特徴量を使用するのではなく，絞り込みを行って静止画像との対応付けを行う．

フレーム画像内において，撮影対象となっている領域以外から抽出した SIFT 特徴量は，Web 上静止画像とのマッチングにおいてノイズとなる．オプティカルフローの利用等によって SIFT 特徴量を抽出する領域を指定すれば，意味的に代表的な特徴点として絞り込むことができる．

ところで，SIFT 特徴量は，画像が小さいほど極小領域の SIFT 特徴量が抽出できなくなるが，少ない演算量で抽出することができる．映像ショットの撮影者は，興味がある部分に注目して撮影すると考えられ，フレーム内において大きい領域であると考えられる．したがって，画像の縮小を行っても，撮影者の興味がある領域からの SIFT 特徴量は抽出できると考えられる．そこで，フレーム画像を縮小することによって，空間的に代表的な特徴点として絞り込む．

また，撮影対象となっているオブジェクトは，オブジェクトの動きやカメラの動きなどがあってもフレームアウトしないように意図して撮影されると考えられる．一方で，撮影者にとって興味の薄い背景領域やオブジェクトは，フレームアウトが生じることがある．このことから，ショット内において短時間ではなく，長時間にわたって連続して撮影された領域は，撮影対象

オブジェクトである可能性が高いと考えられる．ショット内の連続した時刻の 2 つのフレーム画像において，対応付けが可能な局所領域は，フレームアウトしていないと考えることができる．したがって，ショット内の各フレーム画像から抽出できる SIFT 特徴量のうち，フレーム画像間での対応付けが連続して続くものについては，撮影対象オブジェクト領域に含まれる SIFT 特徴量である割合が高いと考えられる．そこで，フレーム画像間で連続して対応付けが可能な特徴点だけを選別することにより，時間的に代表的な特徴点として絞りこむ．類似度の指標としては，特徴点が対応付けられた率を使用し，対応点率＝対応点数 / min(対象特徴点数，静止画像中特徴点数)で定義する．

## 4. 実験と考察

飲料缶を撮影対象として手持ちでの水平移動を含む映像を DV によって撮影し 640[pixel]×480[pixel]とした AVI 画像と，同一のオブジェクトを含み背景が異なる静止画像を用意し，画像縮小，領域の指定，連続出現特徴点による絞り込みをそれぞれ行った場合と，これらを連続して行った場合について，映像ショットから 10 フレーム間隔でサンプリングしたフレーム画像を用いて実験した．画像縮小は，640[pixel]×480[pixel]を基準に，縦横それぞれ 50%，25%とした．領域指定については今回手動で行った．連続出現する特徴点は，可変な W フレームにおいて連続出現する場合について実験した．

画像縮小を行った時の対応点率を図 2 に示す．このとき，元画像からの縮小によって，抽出特徴点数の平均は 238 個→91.1 個→38.8 個と減少したが，対応点率はほぼ減少していなかった．このことから，画像の縮小によって適切に特徴点数の絞り込みを行えたと考えられる．連続出現する特徴点を抽出したときの対応点率を図 3 に示す．連続出現する特徴点による絞り込みにより，特徴点数は減少し，対応点率が向上したことから，適切な領域の特徴点を抽出できたと考えられる．図 4 に画像縮小(25%)，領域指定，連続出現（W=30）による絞り込みと，それらを全て行った時の対応点率を示す．無処理の場合と比較して，高い対応点率をとることから，ショットと類似する静止画像を効率良く判定できたものと考えられる．

## 5. まとめ

ショット内のフレーム画像を縮小すると，抽出できる SIFT 特徴点数が減少するが，本稿では類似度判定に影響を及ぼさないことを示した．その上で，領域指定，連続出現する特徴点の絞り込み後に対応点をとることは，類似度判定に有効であることを示した．ショット全体を対象とした類似度判定を行うことが今後の課題である．

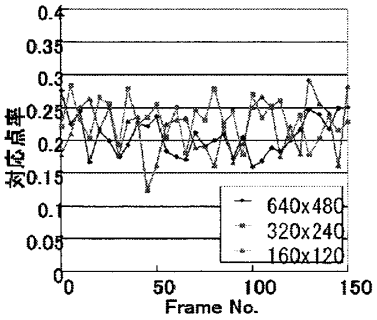


図 2 画像縮小時の対応点率

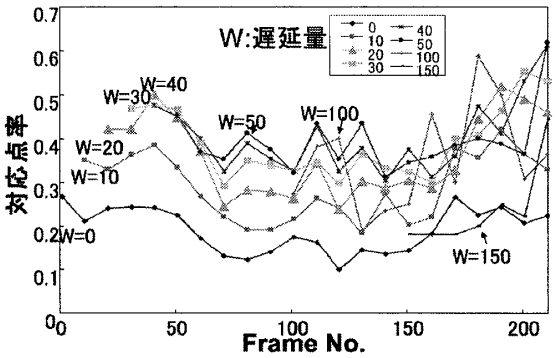


図 3 連続出現特徴点使用時の対応点率

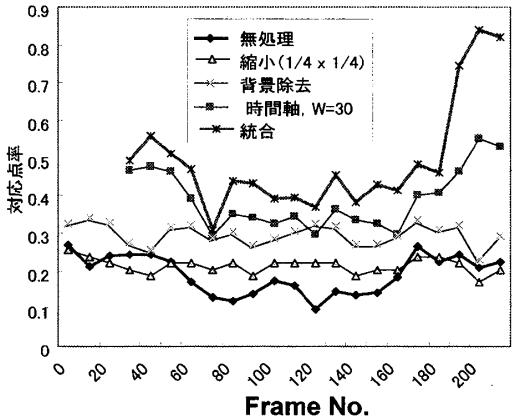


図 4 絞り込みによる対応点率

参考文献

[1] D.G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, IEEE International Journal on Computer Vision, Vol.60, no.2, pp.91-110, 2004

[2] N.Ichimura, “Recognizing Multiple Billboard Advertisements in Videos,”2006, IEEE Pacific-Rim Symposium on Image and Video Technology (PSIVT06), pp.463-473, 2006