I-010

映像メタデータ自動付与実現のための Web 情報を用いた画像マッチング手法の一検討
A Study on a New Image Matching Method using Web Information for Automatic Metadata Generation of Video Contents

関野 真洋† 青木 輝勝‡ 沼澤 潤二‡
Masahiro Sekino† Terumasa Aoki‡ Junji Numazawa‡

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，記憶媒体の発展を背景に，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)．Web 情報を用いるため，映像と静止画像間の対応付けについて検討する．

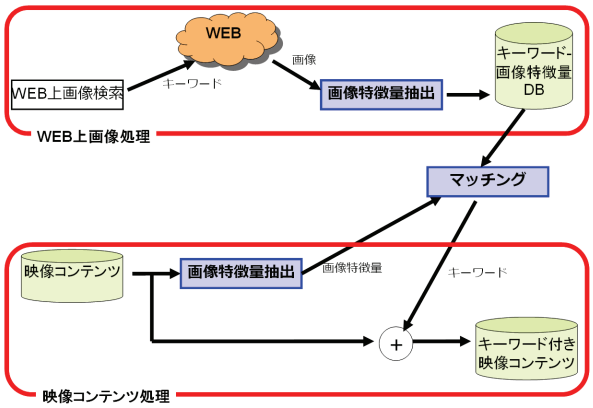


図 1 概念図

2. 画像マッチング

映像コンテンツと Web 上の静止画像との対応付けのため，映像コンテンツ中の最小構成単位であるフレーム画像と，Web 上の静止画像とでマッチングをとる．画像内の特徴点の対応点を探索し，特徴点中対応点が存在した割合が高いときにマッチングしたとする．

画像間の対応点探索に用いる特徴量には，SIFT[1]特徴量を用いる．画像間で類似特徴量を探索し，対応点を決める．

2.1 SIFT と対応点探索

SIFT (Scale Invariant Feature Transform) 特徴量は，Lowe によって提案された輝度勾配に基づく局所特徴量である．SIFT は，画像のスケール変化や回転に不変な特徴量を記述するため，画像のマッチングや物体認識に用いられる．SIFT の処理は，特徴抽出に適した点（以下キーポイント）の検出と，スケール変化・回転・照明変化に不変な特徴量を記述する 2 段階で構成されている．

2.1.1 キーポイント抽出

キーポイント候補は，異なるスケール σ のガウス関数 G(x,y, σ)と入力画像 I(x,y)とを畳み込んだ平滑化画像 L(x,y, σ)の差分をとった画像 D(x,y, σ)から求める． σ の異なる平滑化画像の差分をとることを Difference-of-Gaussian (DoG)と呼ぶ． σ を変え，複数の DoG を得る．

得られた DoG 画像から極値の検出を行う．極値の検出は，DoG 画像 3 枚 1 組で行う．極値の探索を行いたいスケールの画像のある画素に注目したとき，その周辺 8 画素，隣接する上下のスケールの DoG 画像の注目画素と周辺画素 18 画素を比較し，極値であった場合その画素を検出する．検出された画素をキーポイント候補とする．

2.1.2 キーポイント絞込み

検出されたキーポイント候補から，キーポイントとして有効な点を選ぶ．エッジ上のキーポイントを除くため，主曲率を求めて選別する．

2.1.3 輝度勾配方向の割り当て

検出された各キーポイントに対して輝度勾配方向を割り当てる．キーポイントが検出された平滑化画像 L（x,y, σ ）の各画素の勾配 m(x,y)と勾配方向 θ (x,y)を求める．

$$m(x,y) = \sqrt{f_x(x,y)^2 + f_y(x,y)^2} \tag{1}$$

$$\theta(x,y) = \tan^{-1}\frac{f_y(x,y)}{f_x(x,y)} \tag{2}$$

$$\begin{cases} f_x(x,y) = L(x+1,y) - L(x-1,y) \\ f_y(x,y) = L(x,y+1) - L(x,y-1) \end{cases} \tag{3}$$

求めた勾配の大きさ m と勾配の方向 θ から，スケール σ に比例した領域で重み付け方向ヒストグラムを作成する．作成したヒストグラムから最大値となる方向をキーポイントの輝度勾配方向として割り当てる．

2.1.4 特徴量記述

キーポイント周辺領域を 2.1.3 で割り当てた方向を基準とした軸に回転する．この状態で特徴量を算出するため，回転に対する不変性を得られる．分割した周辺領域ごとの各ピクセルの輝度勾配ヒストグラムを作成する．周辺領域を小領域に分割し，それぞれの輝度勾配ヒストグラムを作成する．周辺領域を 4×4 の 16 分割し，それぞれにおいて 8 方向で輝度勾配ヒストグラムを作成する．これによって

† 東北大学大学院情報科学研究科, Graduate School of Information Sciences, Tohoku University
‡ 東北大学電気通信研究所, Research Institute of Electrical Communication, Tohoku University

4×4×8 = 128 次元のベクトルを得る．このベクトルの大きさが 1 となるように正規化し，特徴量とする．

2.1.5 対応点

対応点は次のように求める．

1．まず，各静止画像について，SIFT 特徴量を抽出し，kd 木を構築する．

2．次に，映像中のフレーム画像から SIFT 特徴量を抽出し，キーポイントごとに類似特徴量を作成済みの kd 木より探す．類似特徴量が存在した場合，対応点とする．

3. 提案手法

フレーム画像は，撮影対象の動きや手振れにより，不鮮明であることがあり，不鮮明部分から抽出される SIFT 特徴量は対応点を持たないと考えられる．これをフレーム画像を縮小して解像度を落とすことで軽減する．

そこで，フレーム画像を縮小してから SIFT 特徴量を抽出し，対応点探索を行う．スケールに対して不変な特徴量であるから，入力画像を縮小しても対応点探索は可能であると考えられる．

あるフレーム画像とある静止画像とで対応点探索を行うとき，縮小後のフレーム画像に含まれるキーポイント数を N_F，静止画像に含まれるキーポイント数を N_I，対応点数を N_M としたとき，1 つのキーポイントは複数の対応点を持たないので，$N_M \leqq MIN(N_F, N_I)$ が成り立つ．フレーム画像と静止画像のうち，キーポイント数の少ない画像のキーポイントが対応点を持つ割合が高ければ，対応点探索の効率がよいと考えられるので，以下の式で対応点率を定義する．

$$対応点率 = N_M / MIN(N_F, N_I) \quad (4)$$

フレーム画像，静止画像のどちらか全ての特徴点がもう一方の画像内のいずれかの特徴点と対応付けられたとき 1 となる．

4. 実験と考察

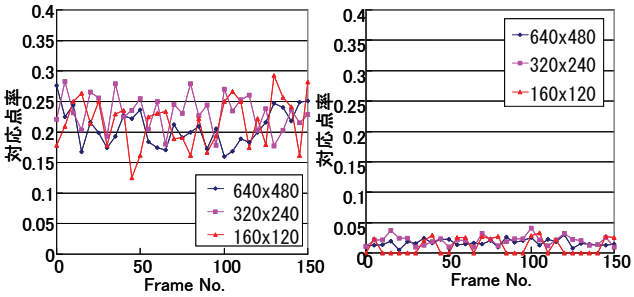
背景を白色とし，実験用映像と静止画像を作成した．映像は DV で撮影し，手持ちでの水平移動を含む．静止画像はデジタルカメラで撮影し，映像中のオブジェクト(以下，対象オブジェクトという)が含まれる場合 5 例と，対象オブジェクト以外のオブジェクトが含まれる場合 5 例とした．

映像，静止画像ともに 640[pixel]×480[pixel]で取り込み，それらを元に 320[pixel]×240[pixel], 160[pixel]×120[pixel]に縮小したものを作成した．

静止画像を一定のサイズとし，映像を縮小したときの対応点探索，映像を一定のサイズとし，静止画像を縮小したときの対応点探索，静止画像，映像ともに同じサイズに縮小して対応点探索を行った．

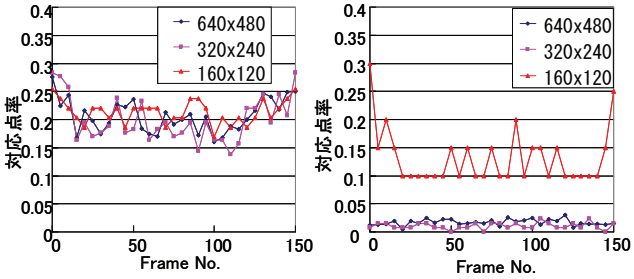
静止画像に対象オブジェクトを含む場合，映像を縮小したときに最も対応点率が向上した．静止画像のみ，映像・静止画像ともに縮小した場合には変化は少ない．静止画像と対応点となるキーポイントは，低解像度となっても抽出できるキーポイントの割合が高かったと言える．

静止画像に対象オブジェクトを含まない場合，画像を 160x120 に縮小したときに急激に対応点率が上昇したが，これは $MIN(N_F, N_I)$が急激に減少したにも関わらず，N_M が変化しなかったためである．低解像度において抽出可能なキーポイントであったと考えられる．静止画像・映像とも に縮小したときの対応点率の上昇フレームも同様の理由である．



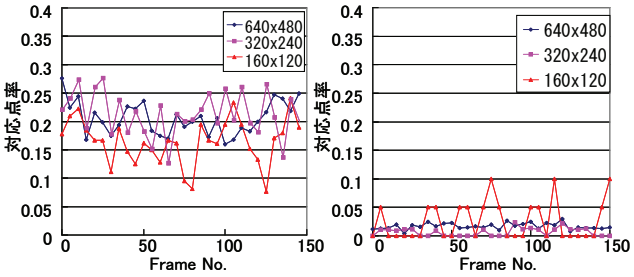
(a)対象オブジェクトを含む画像　(b)対象オブジェクト含まず

図 2 映像を各サイズに縮小したときの映像と静止画像との対応点率変化の一例



(a)対象オブジェクトを含む　(b)対象オブジェクト含まず

図 3 画像を各サイズに縮小したときの映像と静止画像との対応点率変化の一例



(a)対象オブジェクトを含む画像　(b)対象オブジェクト含まず

図 4 映像・画像ともに各サイズに縮小したときの映像と静止画像との対応点率変化の一例

5. まとめ

Web 上静止画像に付与されているメタデータを利用して映像に対してメタデータを付与するため，映像と静止画像との対応付けを行うためのマッチング手法として SIFT を用いた基礎検討を行った．映像中のフレーム画像は動き等により不鮮明な場合があるが，本稿では縮小することで効率よく対応付けを行えることを示した．

参考文献

[1]D. Lowe, “Distinctive image features from scale-invariant keypoints”, Proc. of International Journal of Computer Vision (IJCV), 60(2), pp.91-110, 2004