

画像のグループ化処理を適用した類似画像検索の性能改善

村林 昇† 倉橋 節也† 吉田 健一†

筑波大学大学院 ビジネス科学研究科†

1. はじめに

ビデオカメラが広く普及し、個人でも手軽に旅行や観光などでビデオを撮影して楽しむことが多くなっている。撮影した観光地の風景画像が何であるかという情報は、そのような画像の認識技術が難しいため、現状では記録できるものがない。我々は WWW(World Wide Web) 上には様々な観光や旅行案内に基づいた情報と共に対応する多彩な画像が存在することに着目し、WWW 上における画像と対応する情報を利用してビデオアノテーションへ応用する事を考えた。

ビデオ撮影を考えた場合、必ずしも良好な天候や安定した撮影環境で観光名所を撮影できるとは限らない。このような条件下で撮影されたビデオ画像は、天候の状況による明るさの変化や周囲の陰影などの影響を受け、画質が良好でない場合が多い。またビデオカメラやデジタルカメラの機器間の相違により、撮影画像における色の整合性が取れないこともある。このようなことから、従来の画像検索手法では、ビデオ画像と類似した適切な WWW 画像を十分な精度で検索しビデオアノテーションへ応用することは難しかった。

このような課題に対処するため、グループに基づいた画像検索手法を提案する。この画像検索手法の特徴は、1) wavelet 変換に基づいた画像特徴量を Earth Mover's Distance (EMD) [1] で評価する距離の定義を用いたこと、2) 検索対象の画像を、関係するものどうしグループ化し画像グループ間の類似性に基づいたグループの選択処理を行うこと、の 2 つである。第 1 の特徴は天候の変化など様々な撮影環境下での検索対象の画像間における画質の相違に対応することを目的とし、第 2 の特徴は画像特徴量による検索処理とは別の観点に基づいた検索精度のさらなる改善を目的とする。

2. 関連研究

本研究では、[2]で提案されている画像検索の手法を改良して用いた。[2]で提案されている画像検索における特徴抽出法は、WWW 画像を 25 の領域に分割し領域毎に $L \times U \times V$ 色空間における色特徴を抽出している。そして EMD により画像間距離を求めて WWW 画像における類似画像のクラス分類を行っている。

この画像特徴は画像領域の大局的な特徴であるが、画像の局所特徴を用いることで画像認識を良好な精度で行う手法 (SIFT) が提案されている [3]。本研究では、[2]の色特徴の代わりに wavelet 変換に基づく特徴と EMD を適用した場合の他に、局所特徴量である SIFT 特徴を適用した場合の検索性能についても実験して性能比較を行った。

3. 画像のグループ化処理を適用した画像検索

3.1 wavelet 変換に基づく特徴抽出と EMD

図 1 に示すように、YCbCr 色空間の WWW 画像およびビデオ画像を 240×180 画素の大きさに正規化し、その後 25 領域に分割する。各 25 領域に対して Daubechies-4 wavelet 変換を行い、HH, HL, LH の各成分を $k=15$ とする k-means クラスタリングを行い特徴データに用いる。各クラスのデータ数を重みとして EMD により画像間距離を計算する。

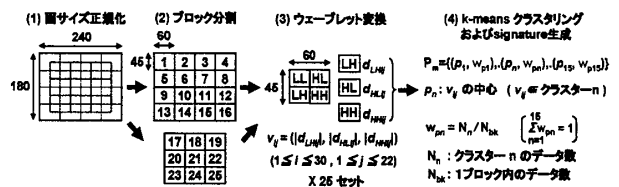


図 1 wavelet 変換に基づく特徴抽出

3.2 グループ処理に基づく画像検索

図 2 において、始めに、ビデオから代表画像を選択する (1)。次に観光情報の WWW ページから

Performance Improvement of Similar Image Retrieval by the Image Grouping Method.

† Noboru Murabayashi (noboru@gssm.otsuka.tsukuba.ac.jp)

† Setsuya Kurahashi (kurahashi@gssm.otsuka.tsukuba.ac.jp)

† Kenichi Yoshida (yoshida@gssm.otsuka.tsukuba.ac.jp)

Graduate School of Business Sciences, University of Tsukuba
(†) Otsuka 3--29--1, Bunkyo-ku, Tokyo, 112--0012, Japan

観光情報と対応する画像グループ集を生成する(2)。選択したビデオ画像とすべての WWW 画像で類似画像検索を行い、グループ内での最類似画像の EMD の総和を計算し、画像グループ距離とする。画像グループ距離が最も小さいグループがビデオ画像に属するとし(3)、そのグループ内で対応する WWW 画像を選択する(4)。

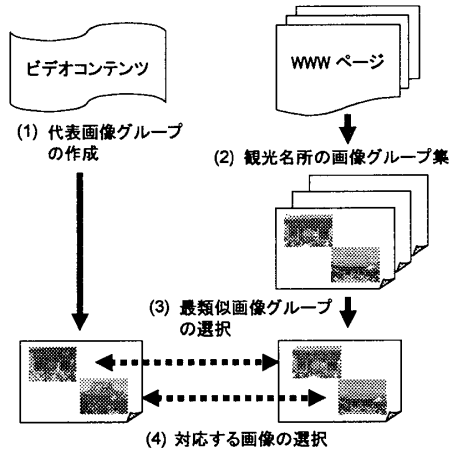


図2 グループに基づいた画像検索法

4. 実験結果

検索性能を検証するため WWW 画像として、幾つかの観光情報の WWW ページから 50 地域を選定した。各地域において 4 枚ずつの画像を選択し、合計 200 枚の画像を用意した。ビデオ画像は上記 50 地域の中から 5 地域を選択し、実際にビデオ撮影した後、代表画像を各地域 4 枚を選択し、合計 20 枚を用意した。

表 1 は、各画像特徴別のグループ正解率とグループ処理を適用した最終の画像正解率を示す。画像特徴のみの正解率は、色特徴は 15% で殆ど性能が出ていない。そのためグループ正解率は 0% で正しく検索できたグループはなかった。この原因として、類似画像同士でもビデオ画像と WWW 画像では撮影条件の相違に基づく色特徴の整合性が取れないことが考えられる。

SIFT 特徴のみの場合は 45% の正解率でグループ正解率は 80% で良好な性能が出ており、最終正解率は 65% であった。Wavelet 特徴のみの場合は、40% の正解率でグループ正解率は 80% で良好な性能であり、最終正解率は 70% で SIFT の場合よりやや良好な性能であった。

この結果から SIFT と wavelet 特徴は最終的な性能の大きな違いは見られないが、図 3 のように画像によっては Keypoint が殆ど検出されず SIFT 特徴では検索性能が出ない画像もあった。

表 1 各画像特徴による実験結果

画像特徴	画像特徴のみ	グループ正解率	最終正解率
色特徴	15 %	0 %	0 %
SIFT	45 %	80 %	65 %
Wavelet	40 %	80 %	70 %

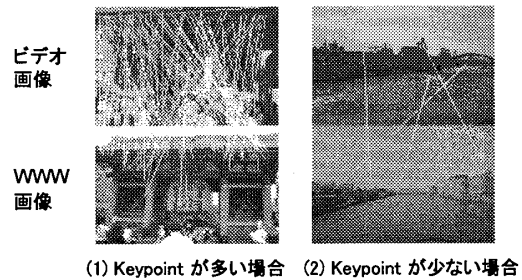


図3 画像による keypoint の相違

5. まとめと今後の課題

wavelet 変換による画像特徴と EMD およびグループに基づく検索法により画像検索性能が改善できることを示した。また局所特徴である SIFT でも wavelet と同程度の検索性能であるが、画像によっては SIFT では検索性能が出ない画像があることが分かった。

今後は、データセット数を増大させた場合における wavelet 変換による特徴と SIFT など局所特徴量とを用いた検索手法の比較、ビデオアニメーションに向けての最適なグループ化方法の検討と性能評価を行う予定である。

参考文献

- [1] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas, "A metric for distributions with applications to image databases", 6th International Conference on Computer Vision, pp59-66, 1998
- [2] 柳井啓司, "一般画像自動分類の実現へ向けた World Wide Web からの画像知識の獲得", 人工知能学会論文誌, vol. 19, No. 5, pp. 429-439, 2004
- [3] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, Vol. 60, No. 2, pp. 91-110 (2004)