

動画像検索のための着目領域発見手法 A method to detect important region for video retrieval

瀧本 政雄[†]
Masao Takimoto

佐藤 真一[‡]
Shin'ichi Satoh

濱田 喬^{†‡}
Takashi Hamada

1. はじめに

近年の動画像の利用の拡大により動画像検索の重要性が増している。そこで本研究では画像データという低次元な情報に基づく動画像処理手法を考案する。低次元情報のみで理想的な動画検索が実現できるとは考えていないが、検索システムの土台としての利用を目標とする。具体的には動画像データに基づいた解析により、動画像の中から着目すべき領域を発見する手法を研究対象とする。このような手法は、例えば人手で意味情報を付加する場合などにも応用が利くものである。

以降、本稿では次章で研究目的とその概要を述べ、3章で目的を実現するための具体的なアプローチを示し、最後に今後の研究予定などをまとめることとする。

2. 研究目的

本研究では画像データに基づく検索を目標としてその基盤とすべき動画像解析を行う。扱う動画像データとしては国立情報学研究所に蓄積されているTV放送の映像データを利用する。TV放送の映像データは内容的にも多岐に富んでおり、また、幅広く利用される可能性があるために汎用的なシステムとしてのテストには適していると考えられる。

単に動画像解析といっても多くの方法があり、本研究では動画像認識に近い処理を研究対象とする。例えば静止画像の検索の前処理として画像認識を利用する研究は広く行われている。しかし完全な画像認識を行うことは困難であるため、本研究では動画像の認識ではなく分類を目標として据える。つまり動画像内に含まれるオブジェクトを認識するのではなく、あるオブジェクトが存在するかどうかの分類、特に動画像内に頻出するような領域が存在する場合はそのような領域に基づいた分類を行う。

そのような目的のために、動画像から着目すべき領域を検出し、その出現を調べることによるショット単位の分類を考えている。着目すべき領域という概念は非常に曖昧なものだが、色合いなどが特徴的な変化を見せる領域は含まれるオブジェクトを判断する上で重要な可能性が高く、さらにその領域がショット内、もしくは動画像データ全体の中で頻出しているならば動画像の内容との関連性も高い領域と考えられ着目領域としてふさわしい。

このように強い特徴を持った領域を自動的に抽出すれば動画像の特徴を示すようなモデルとして扱えるだろう。また、ユーザとの対話を行うことでユーザの意志に従った着目領域発見にも応用できるため、最終的に動画像の領域もしくはオブジェクトに基づいた検索への発展が可能である。その他にも、例えばロゴマークなどが動画像中に頻出しているのかなどといった利用例が考えら

れる。ロゴマークなどは画像的に特徴が強い場合が多く、画像データのみからの抽出も比較的容易であり実現の可能性も高いといえる。

3. 研究のアプローチ

動画像からの領域抽出、そしてそれらの分類や選別は、一般的に静止画または動画の解析法として利用されている手法を組み合わせることで実現可能であると考えている。本章ではそのような手法を紹介し、さらに本研究に適用するにあたっての現状でのアプローチを述べる。

3.1 特徴点抽出による画像認識

一般に静止画像認識の手法は多く提案されているが、ここではWeberらによって提案されている手法[1]に注目した。これは何らかの対象とする物体が含まれる多数の静止画像をトレーニングセットとしてシステムに与えることで、その対象とする物体を含む画像のモデル化を行うものである。そのおおまかな流れをFig.1に示す。

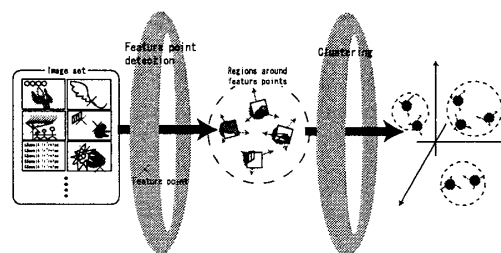


Fig. 1: Weber's image recognition

まず最初に行っている特徴点抽出には多くの検出アルゴリズムが提案されているが、中でもHarris Detection[2]が良い結果を示すという報告が成されており、ここでも類似のアルゴリズムが利用される。

次に得られた特徴点の周囲の領域(11×11)を着目領域として抜き出す。特徴点は強い模様のある箇所主に検出されるため、それを中心に含む領域は着目領域として扱うのに適当である。例えば正面から見た顔の画像の場合、ほとんど目や鼻の穴などが検出されている(Fig.2)。

最後に抜き出した領域のクラスタリングを行う。その結果の内、ある程度の大きさを持つクラスタはトレーニングセットの対象であるオブジェクトの領域を含みやすく、逆に例えば背景などに相当する無関係な領域は小さなクラスタとなるため、分類やモデル化を行える。

この手法を動画像に適用することで動画像における頻出領域が得られることが期待され、本研究の目的の実現に非常に有効であると考えている。例えばショットごとにそのショットを代表するような領域を抽出し、それをさらに全体でクラスタリングすれば結果として動画像全体の特徴をクラスタという形で表せる。

[†]東京大学
[‡]国立情報学研究所



Fig. 2: Feature points

なお、正面向きの顔の含まれる画像をトレーニングセットとして実験を行った。その結果、例えば目を含む領域が多く含まれるような“良い”クラスタも見られたが、全くバラバラの“悪い”クラスタも多かった。これらは小領域の類似度の計算精度の低さやクラスタリングアルゴリズムそのものの限界などが原因である。これらの問題の解決には、小領域の類似度計算法の改善(現状ではカラーヒストグラムや色相関などを利用)や、動画である点をいかした動き情報に基づく領域の事前分類などが有効である。特に後者に関しては次章でとりあげるトラッキングや motion segmentation などの利用を考えている。

3.2 動き情報の利用

効果的な処理のためには抽出した領域の分類や選別を行う必要がある。前節で取り上げた静止画像認識では、クラスタリングがモデルとして利用できる領域の選別に相当する。それに対して本研究では動画像を対象としていることもあり、動き情報を利用することを考えている。そのアプローチとしてトラッキングと motion segmentation をとりあげる。

トラッキング トラッキングはある領域の移動を追跡する手法だが、その例として Tomasi のトラッキング手法 [3] がある。この手法は、連続したフレーム間でトラッキングする上でノイズが最も少ない、つまりトラッキングしやすい領域を抽出して利用する。よって、前節で挙げた特徴点に基づく領域検出は必要ない。逆に着目領域選別の自由度がないという欠点もあるが、実際のところトラッキングしやすい領域とは画像的に特徴がある領域である場合がほとんどで、また、着目領域という定義自体が厳密なものではない。

トラッキングの領域選別への利用法としては、動きによるオブジェクトの分割や、領域の存在期間による重要度の判別などが適当である。実際にはカメラの動きなどによって単純に上のような考え方が適用できない場合も多いが、有望なアプローチといえる。

この Tomasi のトラッキングを走行する車を上から撮影した動画像を対象として動かしたところ、高いトラッキング精度を示すとともにトラッキングを行う領域がオ

ブジェクトを表す上で重要となる領域、つまり着目領域としてふさわしい領域に選ばれていることが確認できた (Fig.3)。車の窓の隅やライトの部分など、オブジェクトを車として判断する上で重要な箇所といえるだろう。

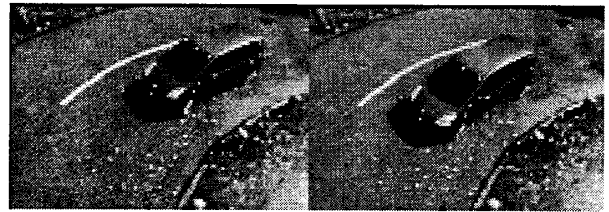


Fig. 3: Tomasi tracking

motion segmentation トラッキングが点や小領域といった画像の一部分に注目した解析手法であるのに対し、motion segmentation はフレーム画像全体を動きにより分割するという手法である。例えば Vasconcelos らによって提案されている手法 [4] では、それぞれの領域の動きをアフィン変換と仮定し、二つのフレーム画像間での各ピクセルの動きから画像を領域分割している。具体的には、分割されるべき各領域はそれぞれがフレーム間での変化を表すアフィン変換のパラメータを持つとして、最も誤差を小さくするようなパラメータと各ピクセルの各領域へ含まれる確率とを推定している。

領域分割によって理想的には動画像をオブジェクトという単位で認識できるようになり、着目領域の分類に非常に効果的であるのは明らかである。実際には理想的な精度での分割は困難であるが、背景と移動物体の分離が可能であるだけでも十分な効果が望めるだろう。

4. まとめ

ここまで研究を行う上での主に方向性を示してきたが、現状ではそれぞれの手法に関した実装や実験を行っている段階である。今後はそれらの手法を上述のような形で組み合わせる上でその結果を検証していく。その他にもユーザとのインタラクションの有無などの検討事項があり、これらを踏まえた上で最終的に検索に応用できるような方式を目指すこととする。

参考文献

- [1] M.Weber, M.Welling and P.Perona. "Unsupervised Learning of Models for Recognition". In *Proceedings on ECCV*, 2001.
- [2] P.Montesinos, V.Gouet and R.Deriche. "Differential Invariants for Color Images". In *Proceedings of 14th International Conference on Pattern Recognition*, Brisbane, Australia, 1998.
- [3] Carlo Tomasi and Takeo Kanade. "Detection and Tracking of Point Features". Carnegie Mellon University Technical Report CMU-CS-91-132, April, 1991.
- [4] Nuno Vasconcelos and Andrew Lippman. "Empirical Bayesian Motion Segmentation". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, February 2001.