

映像内容の重要性を考慮した関心領域推定方式に関する研究

王 国奇[†] 青木 輝勝[‡]東北大学 情報科学研究科[†]東北大学 未来科学技術共同研究センター[‡]

1. はじめに

近年、動画像の中から人間が注目しやすい時空間領域（関心領域）を自動的に抽出する技術について世界中で広く研究を進められている。

人間は視野内から必要と思われる情報を選択し考察する必要がある、興味に従い視覚探索を行う。これら人間の視覚特徴を計算機上で模擬することにより人間が注意を向けやすい画像中の関心領域を推定するための計算モデルがこれまでいくつか提案されてきた。

しかしながら、既存の顕著性マップモデルは2つの問題を抱えている。1つ目の問題は、画像内容の意味を考慮していない（できない）点である。また、2つ目の問題は、既存モデルは多くの場合静止画のみを対象としている点である。動画対応可能な場合にも、具体的な処理方法は動画内の1フレーム毎の処理に帰着しているものが多い。

そこで、本稿では、人間が映像コンテンツを観察するとき、「動きのあるオブジェクトに注意を向ける」という仮説のもと、これらを重要な顕著性要因として利用することにより、特に動きが重要な意味を持つ映像（例：スポーツなど）に対し、既存モデルよりも高精度な顕著性マップを生成できることを実証する。具体的には、オプティカルフローを利用するアプローチによる視線予測技術を提案する。

2. 関連研究及び問題点

ボトムアップ型顕著性マップモデルとして、Itti, Koch らによって提案された顕著度に基づく計算モデル[1]が最も広く知られている。人間の目は一般に色よりも明るさに敏感であることは広く知られていることから、Itti らは第1要素として輝度値を算出している。また、人間は色情報に比較的鈍感であるとは言え、色相環における反対色には強く反応することから、第2要素として色相情報を用いている。そして、人間の目の視覚細胞はある特定の角度（ $\theta =$

$0^\circ, 45^\circ, 90^\circ, 135^\circ$ ）に反応しやすい多段構造となっていることが知られていることから方向性を有するがボールフィルタを用いて人間の視覚機能を模擬している。これが第3要素の方向成分である。

このように Itti-Koch モデルは人間の視覚特徴を計算機上で模擬するモデルとして非常に優れたモデルである。しかしながら、このモデルはあくまでボトムアップ型視覚特徴を模擬したに過ぎず、また、静止画に基づく処理であるため、通常の「何らかの意味を持つ」映像コンテンツに対しては生成された顕著性マップの予測精度が高くできない問題がある。

3. 動き情報に基づき顕著性マップ生成方式の提案

前述した既存モデルの問題点を解決するため、筆者らは、人間が映像コンテンツを観察する時、「動きのあるオブジェクトに注意を向ける」という仮説のもと、これらを重要な顕著性要因として利用することにより、スポーツ映像やドラマなどのように動きが重要な意味を持つ映像に対し、既存モデルよりも高精度な顕著性マップを生成可能なモデルを提案する。

提案方式は Itti-Koch モデルをベースとして、そこに動き情報を加えたものである。提案方式のアーキテクチャを図1に示す。

図1の処理の流れは以下の通りである。

(1)入力画像に対して、Itti らのモデルに必要な色相情報・輝度値・方向成分に加え動き成分を求める。

(2)特徴マップ (Feature maps) を作成するとき、同時に動き成分の特性マップも生成する。

(3)目立ち度マップ (Conspicuity maps) を作成するとき、同時に動き成分の目立ち度マップも生成する。

(4)色相情報・輝度値・方向成分の以外に、動き成分を合わせて顕著性マップを生成する。

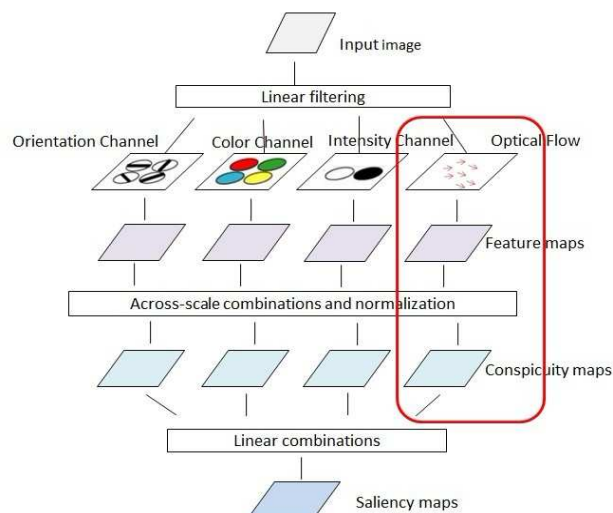


図1. 動き情報を入れる Itti モデルのアーキテクチャ

4. 動き情報に基づく特徴マップ、目立ち度マップの生成と評価

提案方式では、動き情報に基づく特徴マップ、目立ち度マップを生成するためにオプティカルフローを用いることとした。

オプティカルフローを計算する手法として、大別してブロックマッチング法（BM(Block Matching)法など）と勾配法（LK(Lucas-Kanade)法、ピラミッドLK法など）が存在する。今回の実験で筆者らはノイズ、誤差や計算時間などを考慮した上で、密なオプティカルフローを計算でき精度の良い Gunnar Farneback 法を採用した。

オプティカルフローにおいて、動きベクトルの長さはオブジェクトの動きの程度を表している。そこで、動きベクトルは長ければ長いほど、オブジェクトの動きが激しく、人間の注意を向けるという仮説のもと、動きベクトルの長さにより 0~255 に正規化された輝度値を各画素に設定した顕著性マップを生成する(図2)。

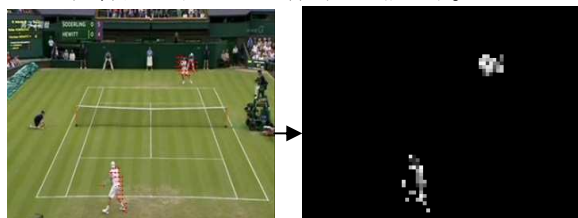


図2. オプティカルフロー画像による生成した顕著性マップ

次に、提案した動き情報を用いた顕著性マップと Itti らのボトムアップ型の注意の顕著性マップの比較を行う。このため、視線追跡装置を用いた視線計測実験を行い、ビデオを観賞時の

視線位置を取得する。この結果は、上述の2つの顕著性マップの精度を比較する際の模範解答として用いることができる。

図3に、スポーツ（テニス）に対して評価実験を行った結果を示す。図3は、視線計測実験結果(A)、Itti-Koch モデルに基づく顕著性マップ結果(B)、筆者らが提案した動き情報で重要度を表す顕著性マップ結果(C)を生成した結果を示している。この図3から明らかなように、(A)と(C)の顕著性マップは非常によく似ているが、(A)と(B)のマップは必ずしも似ていない。上述の通り、(A)は模範解答に相当するマップであり、(A)と似ているほど精度の高いマップと考えられることから、少なくとも実験で用いたスポーツ（テニス）の動画に対しては、Itti-Koch モデルに基づく顕著性マップ(B)よりも筆者らの提案方式(C)のほうが高精度であると言える。

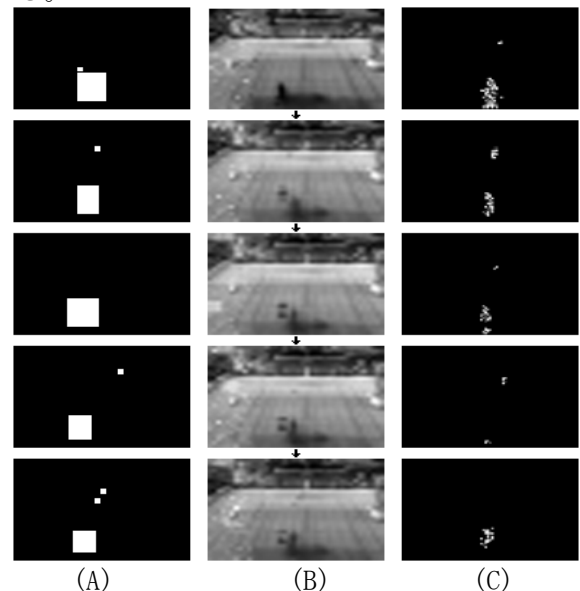


図3. 実験結果

5. まとめ

本稿では、動き情報を考慮した顕著性マップの生成方式を提案し、スポーツのシーンを対象とした実証実験を行った。その結果、今回提唱した動き情報顕著性マップに対して、ある程度の有効性を示すことができた。

文 献

- [1] L. Itti, C. Koch, and E. Niebur, "A Modal of Saliency-Based Visual Attention for Rapid Scene Analysis," Proc. Of IEEE Transactions of Pattern Analysis and Machine Intelligence, Vol. 20, No. 11, pp. 1254-1259(1998).