

運転手注視点予測のための混合エキスパート型モデルを用いた
動的顕著性マップ構築法
Construction of Dynamic Saliency Maps to Predict Drivers' Gazes
via a Mixture of Experts-Type Model

中澤 空知[†] 中田 洋平[‡]
Sorachi Nakazawa Yohei Nakada

1. はじめに

近年、運転支援技術は著しい発展を見せており、自動車製造会社のみならず、様々な企業や機関で研究開発がなされてきている。このような背景の下、本論文では、運転時における人の目の行きやすさを数量化する動的顕著性マップの構成法を提案する。

提案手法では、車載カメラで撮影された入力画像から複数の特徴マップを算出する。そして、この複数の特徴マップに対して、センターバイアス処理と正規化処理を施してから組み合わせることで、動的顕著性マップを構成する。また、各正規化後特徴マップの組み合わせ比率を、入力画像から得られる運転状況を表す運転状況特徴量をもとに、混合エキスパート型のモデルにより動的に決定する。このような動的なモデルにより、運転状況に応じた動的顕著性マップが構成されることが期待できる。また、提案手法の検証を目的とし、アイマークカメラ動画と車載カメラ動画を含む DR(eye)VE データセット[1]から整備した注視点データを用いた実験を実施する。

2. 車載カメラ画像に対する注視点データの整備

本研究では、DR(eye)VE データセット[1]から、車載カメラ動画に対する注視点データを整備している。同データセットは、学習用と評価用に区分けされており、双方とも 37 シーンの合計 74 シーンで構成されている。また、各シーンは、300 秒である。各シーンに対して、実際の運転環境で撮影された車載動画に加え、アイマークカメラ動画、注視点座標値、運転時の天候、道路環境、速度などの情報が与えられている。

ただし、注視点座標値は、車載動画でなく、アイマークカメラ動画に対応しており、そのため、提案手法を検証するために、アイマークカメラ画像上の注視点データを車載カメラ画像上に空間的な対応付けを行う必要がある。本研究では、特徴点マッチングに用いられる SIFT, SURF, AKAZE から得られた 3 種類の特徴点を用いて、アイマークカメラ画像と車載カメラ画像の 2 つの画像で空間的対応付けを行い、それによって注視点データを整備している。

3. 提案手法

前述のように提案手法で構成される動的顕著性マップでは、複数の特徴マップを組み合わせる際、運転状況を表す画像特徴量（運転状況特徴量）により組み合わせ比率を動的に決定する。

3.1 特徴マップ

視覚的顕著性マップの先行研究である Itti モデル[2]や BMS モデル[3][4]に基づいて算出した 4 つの静的な特徴マップと、それらの静的な特徴マップと入力画像から算出したオブティカルフローの大きさに基づく 5 つの動的な特徴マップから計 9 つの視覚的顕著性マップを構成する。また、各画素値に最低限のバイアスを与えてその確率値を平滑化するために、一様な画素値を持つ一様特徴マップを 10 番目の特徴マップとして用いている。さらに、運転時には、人間の視線は消失点に引きつけられやすいこと[1]が知られており、その性質を反映させるために、10 番目以外の特徴マップには、センターバイアス処理を施している。

3.2 運転状況特徴量

本研究では、運転状況を表す特徴量として、入力画像から得られる色情報に関する統計量 10 種類とオブティカルフローに関する統計量 8 種類を用いる。なお、基本的には、色情報に基づく統計量は運転場所に関する情報を、オブティカルフローに基づく統計量は自動車の移動に関する情報を捉えるためのものとなる。ただし、後述の検証実験においては、これらの運転状況特徴量を実際に用いる際には、学習用データを用いた主成分分析により、18 次元から 4 次元に次元削減している。

3.3 混合エキスパート型モデル

提案手法では、動的顕著性マップを注視点の存在確率を表す確率分布として捉える。そして、センターバイアス処理済みの正規化後の特徴マップを要素分布とする混合分布として表す。さらに、混合エキスパート[5]の考え方を導入することで、前述の運転状況特徴量（ベクトル） ξ を入力とする関数を組み合わせ比率としている。すなわち、動的顕著性マップは、次式に表すような条件付きの混合分布として表される。

$$p(x, y | \xi) = \sum_{k=1}^F g_k(\xi) f_k(x, y) \quad (1)$$

ただし、ベクトル (x, y) は画像上での注視点の座標ベクトルを表し、関数 $f_k(x, y)$ は k 番目 ($k = 1, 2, \dots, F$) のセンターバイアス処理済みの正規化後特徴マップを表している。自然数 F は特徴マップの数であり、本研究では $F = 10$ となる。関数 $g_k(\xi)$ はゲート関数とも呼ばれ、任意の運転状況特徴量 $\xi \in \Xi$ で、次式を満たすものとなる。

$$\sum_{k=1}^F g_k(\xi) = 1, \quad 0 \leq g_k(\xi) \leq 1 \quad (k = 1, 2, \dots, F) \quad (2)$$

なお、 Ξ は運転状況特徴量 ξ の範囲を表す集合である。

また、提案手法では、ゲート関数 $g_k(\xi)$ は次式のように

[†] 明治大学大学院 先端数理科学研究科
Graduate School of Advanced Mathematical Sciences, Meiji University

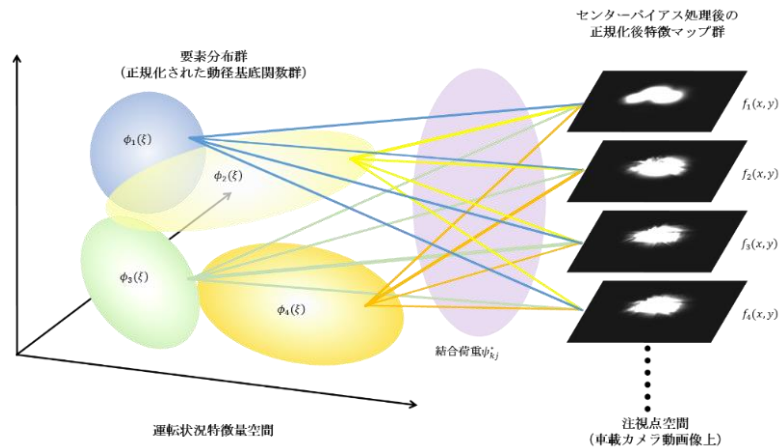


図1 ゲート関数の仕組みの模式図

表1 評価用データに対する平均対数尤度値の比較

モデル	平均対数尤度	標準誤差
提案手法	-11.0913	0.0167
BMS モデル	-13.0125	0.0077
Itti モデル	-13.4591	0.0064
BMS モデル (センターバイアス処理後)	-11.1726	0.0236
Itti モデル (センターバイアス処理後)	-11.3244	0.0226
センターバイアス	-11.2540	0.0209

書き表すことができる。

$$g_k(\xi) = \frac{w_k^*(\xi)}{\sum_{k'=1}^F w_{k'}^*(\xi)} \quad (3)$$

ただし、関数 $w_k^*(\xi)$ は次式のように、運転状況特徴量空間上の混合多変量正規分布としている。

$$w_k^*(\xi) = \sum_{j=1}^M \psi_{kj}^* \phi_j(\xi) \quad (4)$$

ここで、関数 $\phi_j(\xi)$ は j 番目($j = 1, 2, \dots, M$)の要素分布を表し、本研究では多変量正規分布である。また、自然数 M は要素分布数を表す。

このような設定の下では、ゲート関数 $g_k(\xi)$ には、動径基底関数ネットワーク型の構造が浮かび上がる。より具体的に述べれば、運転状況特徴量空間上には、確率的に分割するための要素分布である多変量正規分布 $\phi_j(\xi)$ が置かれている。それらを(非負の)結合荷重 ψ_{kj}^* により組み合わせた F 個の出力を式(4)により求める。その後、これらの出力を、式(3)のように和が1となるように正規化する。そして、正規化された出力を正規化後特徴マップの組み合わせ比率として、動的顕著性マップを算出する構造が見て取れる。図1には、このようなゲート関数の仕組みの模式図を示す。

なお、要素分布 $\phi_j(\xi)$ については、そのパラメータと要素数 M は、事前に運転状況特徴量上で多変量混合正規分布の学習することで決定している。また、組み合わせ比率

(結合荷重) ψ_{kj}^* については、要素分布 $\phi_j(\xi)$ が固定された下で、EMアルゴリズム[5]を用いて学習している。

4. 検証実験

DR(eye)VE データセット[1]を用いた検証実験においては、前述の方法にて、そこから学習用データと評価用データを作成している。ただし、その際に、各シーンを1秒間隔でサンプリングしている。また、結合荷重や要素分布の学習においては、学習用データのみを用いて学習を行っている。表1は、提案手法で構成した動的顕著性マップを含めたいくつかの視覚的顕著性マップに対する評価用データでの平均対数尤度値の比較結果を示している。この値が高ければ高いほど、確率分布としての意味において注視点の予測性能が高いことを表している。同表から、本実験においては提案手法が比較的良好に機能していることが分かる。

5. 今後の課題

今後は、より有効な特徴マップや運転状況特徴量を導入したり、センターバイアスを動的にしたりすることなどで、更なる高精度化を図っていく。

謝辞

同研究室に所属していた牛島郷瑠さんには、本研究を進めるにあたりご議論頂きました。感謝いたします。また、DR(eye)VE データのダウンロードの際に、ご助力頂きました AImageLab の皆様に感謝いたします。

参考文献

- [1] AImageLab, DR(eye)VE Project, <http://aimagelab.ing.unimore.it/dreyeve>
- [2] L. Itti, C. Koch, E. Niebur, "A Model of Saliency-based Visual Attention for Rapid Scene Analysis", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 20, No. 11, pp.1254-1259 (1998).
- [3] J. Zhang, S. Sclaroff, "Saliency Detection: A Boolean Map Approach", Proc. of IEEE International Conference on Computer Vision 2013 (ICCV2013), pp. 153-160 (2013).
- [4] J. Zhang, S. Sclaroff, "Exploiting Surroundedness for Saliency Detection: a Boolean Map Approach", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 38, No. 5, pp. 889-902 (2016).
- [5] C. M. ビショップ, パターン認識と機械学習 上・下, 丸善出版 (2012).