

視線停留発生時刻の分散を用いたユーザ所望画像識別の検討

Classification of User's Desired Images Using Dispersion of Fixation Occurrence

西口 侑希†
Yuki Nishiguchi

菅沼 睦‡
Mutsumi Suganuma

亀山 渉*
Wataru Kameyama

1. まえがき

近年、画像検索・共有サイトの発達や大容量記憶媒体の低価格化などにより、膨大な画像データを個人で扱うことが容易になってきている。これに伴い、画像データの効率的な検索方法の需要が高まっている。

コンピュータに画像の意味内容を理解させる技術はまだ確立されていない。そのため、画像データの内容をメタデータとして表し、画像に付与することが一般的である。しかし、メタデータには、付与コスト、アノテーションの個人依存、内容を言語表現することの限界などの問題が存在する。

そこで、ユーザが検索システムに対してフィードバックを返す適合性フィードバック[1]が注目されている。画像検索システムに対しては、ユーザの視線情報をフィードバックとして利用し、類似画像を視線情報で識別する試み[2,3]や画像特徴量と視線情報を用いて画像の類似度ランキングを推測する試み[4]がなされている。

視線情報によるフィードバックは、クリックなどのフィードバックとは異なり、画像閲覧以外の余分な動作を必要としない。また、視線情報はユーザに固有の情報であるため、個人に適応した検索システムを実現できる可能性がある。

このような適合フィードバックに基づく画像識別性能向上のため、筆者らは、瞳孔径に基づく CPR 特徴量を提案している[5]。本報告では、これに加えて視線停留発生時刻の分散を特徴量として追加することで、さらなる識別性能の向上を試みる。

2. 実験

2.1 実験概要

所望画像を探すときの視線情報と瞳孔径を収集するため、実験を行った。実験では、被験者に所望画像としてクエリを与え、クエリに類似する画像を探すタスクを課すことで、所望画像探索時の視線情報と瞳孔径を擬似的に収集できるものと考えた。

2.2 使用機器

視線情報と瞳孔径は Tobii X60 Eye Tracker を用いて記録した。視線停留の検出には OGAMA[6]を用いた。画像は NANA0 23 インチ LCD モニター (FS2332) に提示した。

2.3 被験者

被験者は年齢が 22 歳から 30 歳の 14 人 (男性 13 人, 女性 1 人) から構成され、平均年齢は 25.14 歳、年齢の標準偏差は 2.18 歳であった。

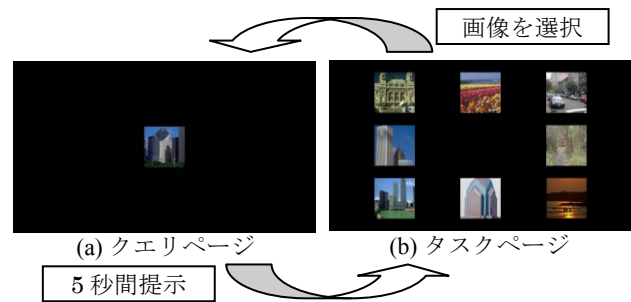


図 1 実験タスク概要

2.4 被験者タスク

図 1 に実験タスク概要を示す。被験者には、クエリ画像 1 枚からなるクエリページと 8 枚のタスク画像群からなるタスクページを提示した。クエリページを 5 秒間提示した後、タスクページを表示し、画像群からクエリ画像に最も類似していると思う画像を一枚キーボードで回答させた。クエリページの提示から回答までを 1 セットとし、これを 26 セット繰り返した。ただし、先頭 5 セットは練習試行とみなして、識別には用いなかった。

使用した画像は LabelMe 8 Scene Categories Dataset[7] である。正解カテゴリは“tall building”とし、識別に用いた 1 被験者当たり 168 枚のうち 63 枚を正解画像とした。なお、クエリは実験を通して 1 種類のみが提示された。

3. 提案手法

3.1 視線停留発生時刻の分散

視線停留発生時刻の分散を式(1)のように算出する。

$$dispersion = \frac{1}{n} \sum_{i=1}^n (x_i - m)^2 \quad (1)$$

ここで、 n は視線停留の回数、 x は視線停留発生時刻、 m は視線停留発生時刻の平均である。視線停留発生時刻は各タスクページでの滞在時間で正規化する。また、 n が 2 未満のとき、分散は 0 とする。

4. 評価

4.1 対光反射補正

瞳孔径は認知活動に影響を受けることが報告されており[8]、筆者らは瞳孔径に基づく CPR 特徴量を提案している[5]。しかし、瞳孔径は認知活動のみだけでなく、輝度変動にも影響を受ける。そのため、輝度変動のみに影響された瞳孔径を予測する必要があることから、対光反射補正を行った。対光反射補正は文献[9]に基づいて試みたが、補正を行った場合に識別性能が低下したため、瞳孔径予測モデルの精度が不十分であると考えた[3]。した

†早稲田大学大学院国際情報通信研究科

‡早稲田大学国際情報通信研究センター

*早稲田大学基幹理工学部

がって、本報告では対光反射補正せずに算出した CPR 特徴量を用いている。

4.2 特徴量の比較

本報告で識別に用いた計 14 次元の特徴量を表 1 に示す。提案特徴量の有効性を確認するため、(i) No.1~11 の 11 次元特徴量、(ii) CPR 特徴量を追加した 13 次元特徴量 (No.1~13)、(iii) 分散特徴量を追加した 14 次元特徴量 (No.1~14) で識別性能を比較した。識別には、サポートベクターマシン (SVM) の正接カーネルを用いて、14 人分のデータで 14 分割交差検証を行った。

5. 結果と考察

図 2 に各手法に対する ROC 曲線を示す。図 2 より分散特徴量を追加することで、識別性能が向上することがわかった。

文献[10]では視線停留発生時刻の標準偏差を特徴量として用いているが、効果的な特徴量であるとは報告されていない。しかし、本報告で行った実験環境では、識別に有効な特徴量であると考えられる。

表 1 識別に用いた特徴量

No.	特徴量	出典
1	対象画像内の総視線停留回数	[3],[4]
2	対象画像内の総視線停留時間	[2],[3],[4]
3	対象画像内の平均視線停留時間	[2],[3],[4]
4	対象画像内の最大視線停留時間	[2]
5	対象画像内の初回視線停留時間	[2],[4]
6	No.2 のページ内順位 ("1","2",..., "8")	[2]
7	No.3 のページ内順位 ("1","2",..., "8")	[2]
8	No.4 のページ内順位 ("1","2",..., "8")	[2]
9	総視線停留時間/ページ滞在時間	[2]
10	平均視線停留時間/ページ滞在時間	[2]
11	最大視線停留時間/ページ滞在時間	[2]
12	積算瞳孔径(CPR)最大値	[5]
13	CPR 平均値	[5]
14	視線停留発生時刻の分散	

6. まとめと今後の課題

本報告では、視線停留発生時刻の分散を特徴量として用いることで、視線情報による類似画像識別性能の向上を確認した。

今後は個人に適応した画像識別・分類を目指し、個人内予測を行う必要がある。また、画像に対する注意の向け方は個人によって異なることが考えられるが、本報告では、複数の被験者によるデータで識別器を構築し個人間識別を行っているため、個人間で特徴量の値に大きく差が出てしまい、効果的に作用しなかった可能性がある。

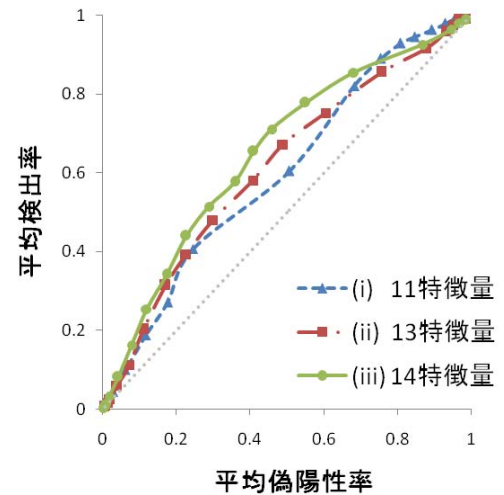


図 2 各特徴量に対する ROC 曲線

さらに、CPR 特徴量を用いるにあたって、対光反射補正を行うために瞳孔径予測モデルの精度向上も今後の課題としたい。

謝辞

本研究は JSPS 科研費 25330136 の助成による。

参考文献

- [1] T. S. Huang, "Active Learning for Interactive Multimedia Retrieval," *Proceedings of the IEEE*, 96, 4, pp.648-667, 2008.
- [2] S. N. Hajimirza, M. Proulx, and E. Izquierdo, "Reading Users' Minds From Their Eyes: A Method for Implicit Image Annotation," *IEEE Trans. Multimedia*, 14, 3, pp.805-815, 2011.
- [3] A. Klami, C. Saunders, T. E. de Campos, and S. Kaski, "Can Relevance of Images Be Inferred from Eye Movements?," *MIR '08 Proceedings of the 1st ACM international conference on Multimedia information retrieval*, pp.134-140, 2008.
- [4] K. Pasupa, Craig J. Saunders, S. Szedmak, A. Klami, S. Kaski, S. R. Gunn, "Learning to Rank Images from Eye movements," *IEEE 12th International Conference on Computer Vision Workshops*, pp.2009-2016, 2009.
- [5] 西口侑希, 菅沼睦, 亀山渉, "視線情報による画像識別に向けた瞳孔径特徴量の提案," *電子情報通信学会, 2014 年総合大会, D-12A*, 2014.
- [6] <http://www.ogama.net/> (last visited on 2014-6-20)
- [7] <http://people.csail.mit.edu/torr/alba/code/spatialenvelope/> (last visited on 2014-6-20)
- [8] T. Partala and V. Surakka, "Pupil size variation as an indication of affective processing," *Int. J. Human-Computer Studies*, 59, pp.185-198, 2003.
- [9] 浅野樹美, 安池一貴, 中山実, 清水康敬, "輝度変化に対する瞳孔面積変化モデル," *電子情報通信学会論文誌(A)*, Vol.J77-A, No.5, pp.794-801, 1994.
- [10] L. Kozma, A. Klami and S. Kaski, "GaZIR: Gaze-based Zooming Interface for Image Retrieval," *Proceedings of the 2009 international conference on Multimodal interfaces*, pp.305-312, 2009.