

#### **NTT** Communication Science Laboratories

## 【サーベイ論文】 人間の視覚的注意の計算モデル

#### 木村昭悟(1) 米谷竜(2) 平山高嗣(3)

- (I) 日本電信電話(株) コミュニケーション科学基礎研究所
- (2) 京都大学 大学院情報学研究科
- (3) 名古屋大学大学院情報科学研究科



#### こんなふとした思いつきからだった



Akisato Kimura @ akisato

これを機会に関連分野のレビューでもしたいなぁ…賛同 者ゆるぼ / My latest SlideShare upload : 人間の視覚 的注意を予測するモデル – 動的ベイジアンネットワー クに基づく最新のアプローチ – slidesha.re/rPG2Mv



2011年12月27日 - 11:08 Tweet Buttonから、このツイートをサイトに埋め込む

🛧 返信 👜 削除 🌟 お気に入りに登録

### そして1時間半後...

釣れた!!



Akisato Kimura

じゃ,やるか.3月PRMUに向けて…??RT @yonetaniryo:@\_akisato 視覚的注意モデルは,D 論書くにあたって,どこかのタイミングである程度まとまっ たサーベイしておかないといけないのです…

2011年12月27日 - 12:55 HootSuiteから、このツイートをサイトに埋め込む

🛧 返信 👜 削除 🌟 お気に入りに登録



#### • 引き返せなくなりました...



Akisato Kimura @ akisato

思いついてわずか1日, ほとんど 勢いで 申し込んでしまった. しかし, 申し込んだからには, きちんとしたものを 作るぞ! **#prmu** 



2011年12月28日 - 18:51 HootSuiteから、このツイートをサイトに埋め込む

🗲 返信 🏛 削除 🌟 お気に入りに登録

## 何を目指すのか?

- 聴講においでの皆様へ
  - 「人間の視覚的注意の計算モデル」は、
     PRMUとHIPの境界領域にあるテーマ.
  - •相互理解を図る上で一つの切り口になれば.
- (主に) 我々自身にとって
  - 関連研究成果を外部に発信する際に,
     既存研究の系統だった深い調査分析が必須.

おしながき

#### 1. 視覚的注意,って何ですか?

- 2. 人間の視覚的注意,って, どこまで解明されているんだろう?
- 3. 人間の視覚的注意を真似する方法
  - a. 画像しか使わないで真似をする
  - b. 画像以外の情報も用いて真似をする
- 4. これから何をすると面白いか?

## 視覚的注意とは何か?

#### 視覚的注意を体感してみよう



・目立つものに
目が向きやすい
・人間の顔には
目が向きやすい
・真ん中(最初の 視線位置)の近くに
目が向きやすい
・判断はほぼ瞬時

## 視覚的注意とは?

 目から入力されてくる信号の中から 重要と思われる情報を瞬時に判断し, 選択的に情報を獲得するための機構

- 画像中のどの部分が重要か瞬時に判断?
   → 計算機に実装できたらすごいかも?!
- 「人間の視覚的注意の計算モデル」
   研究の発想と動機はここにあり

#### そうか,重要な領域を取れば…

- 視覚的注意を実装できた……ではない!!
  - 重要領域の抽出と、視覚特性のモデル化を、
     混同してはいけない!
  - 人間の視覚特性のモデル化が重要なのか, 重要領域の抽出が重要なのか, その区別ができていない研究が散見される.
  - この状況を何とか是正しないと!
     (実は、今回のサーベイの最初の動機はここ、)

### 重要領域抽出の方法を分類してみる



# とある計算モデルを使ってみると



## 視覚的注意について どこまで解明されているか?

## 視覚探索 という課題があります

目標刺激(見つけたいもの)を
 多数の妨害刺激(不要なもの)から
 被験者に見つけてもらう課題.



(Cf. http://www.l.u-tokyo.ac.jp/AandC/HLV/DataBase/VS/intro.html)

#### 視覚探索が解明の鍵を握る

# 目標刺激と妨害刺激の関係を適切に操作 → 視覚処理の構造が見えてくる



(Cf. http://www.l.u-tokyo.ac.jp/AandC/HLV/DataBase/VS/intro.html)

#### *特徵統合理論* [Treisman+1980]





- •特徴探索条件では、並列処理が可能.
- 結合探索条件では、必然的に逐次処理.



#### 特徴統合理論を軸に研究が進む

- 特徴統合理論では説明できない事象を どのように説明するか?が研究の種.
  - 同じ視覚刺激でも課題の与え方一つで 処理時間が大きく変わる.
     [Nakayama 1990] [Bravo+ 1992]
     → トップダウン制御が関わる問題.
  - 視覚刺激間の類似性が探索の難易度に影響.
     [Duncan+1989,1989,1992]
     、 あとで出てきます







 結合探索条件の視覚探索において, 探索対象の特性を陽に組み込んでいる.



#### • 与えられたタスクによる変動 [Yarbus 1967]



-ベイ: 人間の視覚的注意の計算モデル

#### 与えられたタスクによる変動 [Yarbus 1967]



#### 絵画中の人間の衣服を 覚えて下さい.

ベイ: 人間の視覚的注意の計算モデル

#### • 与えられたタスクによる変動 [Yarbus 1967]



#### 絵画中の物体の配置を 覚えて下さい.

- ベイ: 人間の視覚的注意の計算モデル

#### 与えられたタスクによる変動 [Yarbus 1967]



#### 自由に観察して下さい。

-ベイ: 人間の視覚的注意の計算モデル

# 同一画像に対する視線運動は 必ずしも同一ではない [Yarbus 1967]



#### 心理状態などに起因する 不確定性の可能性

純粋な確率的な揺らぎ?

ベイ: 人間の視覚的注意の計算モデル

#### まとめると

- 単一特徴での処理は並列処理.
- 複数特徴での処理の多くは逐次処理.
- 探索対象の特性が既知の場合,
   その特性を強調する処理が行われる.
- 意図や認知心理状態だけでなく,
   いくつかの要因で注意位置が変動する.

# ボトムアップ計算モデル (画像だけを使うモデル)

#### Saliency Map model [Itti+ 1998]

- 特徴統合理論の流れを汲む計算モデル
  - 多重解像度モデル [Koch+Ullman 1985] を 計算機に実装可能な形としたモデル
- 実質的なdefacto-standard



イ: 人間の視覚的注意の計算モデル

## Saliency Map modelの実装





多重解像度モデルを 考慮

結合探索条件での 逐次処理は、出力後の Inhibition of Returnで実現

## Saliency Map modelの発展 1

- Gradual depletion model [Leung+ 2007]
  - Neural adaptation [Hartline 1940] を実装



## Saliency Map modelの発展 2

Stereo saliency map [Jeong+ 2008]



## そういえば

- 特徴統合理論では説明できない事象を どのように説明するか?が研究の種.
  - 同じ刺激を見せても,課題の与え方一つで 処理時間が大きく変わる.[13][14]
     → トップダウンの影響
  - 目標刺激と妨害刺激の類似性, 妨害刺激間の類似性が探索の難易度に影響.
     → あとで出てきます.

     <u>まだ解決していない!!!</u>

#### *信号検出理論* [Eckstein+ 2000]

#### • 特徴統合理論に確率的なゆらぎを追加



## 信号検出理論を導入したモデル 1

• 確率的誘導探索モデル [Koike+2002,2006]



## 信号検出理論を導入したモデル 1

#### • 動的ベイジアンネットワーク [Pang+2008]



#### 時間的顕著性の考慮

事後確率変わらず Bayesian surprise P(M D) [ltti+ 2009] [Baldi+ 2010] posterior MTV OWN FOX BBC Not very surprising Not very surprising 次に起こりそうな P(M) イベントを予測, prior 事前確率として保持 MTV ONN FOX BBC Snow Very surprising Very surprising P(M | D) posterio ここが詳しいです:pooneilの脳科学論文コメント MTV CNN FOX BBC Snow http://pooneil.sakura.ne.jp/archives/permalink/000964.php 事後確率変動! サーベイ: 人間の視覚的注意の計算モデル 36 (Cf. http://ilab.usc.edu/surprise/)
## 人間の視覚特性と関係ないモデル

- 信号処理・CV・MLでの主流
- 信号の非正規性・非定常性に立脚
  - Spectral residual
  - 平滑化差分
- 対象非依存の事前知識を利用
  - 基底との自己情報量
  - Ad-hocなマルチモーダル特徴量の組合せ

# 信号の非定常性に基づくモデル 1

#### • Spectral residual [Hou+ 2007]

自然界に幅広く見られる1/fゆらぎに着目.



# 信号の非定常性に基づくモデル 2

- 平滑化差分方式 [Achanta+ 2008,2009]
  - Itti modelをLab表色系版 → さらに簡略化



(Cf. http://ivrgwww.epfl.ch/supplementary\_material/RK\_CVPR09/index.html)

## 事前知識に基づくモデル 1

#### • ICA基底との自己情報量 [Bruce+ 2006]



## 事前知識に基づくモデル 2

• 使えるものは何でも使う [Ma+2003,2005]



## 事前知識に基づくモデル 2

● 使えるものは何でも使う [Ma+ 2003,2005]



# トップダウン計算モデル (画像以外も使うモデル)

トップダウン要素による分類

(1)探索目標の事前学習型(2)文脈依存型(視覚探索以外のタスク,状態/状況依存)

技術要素による分類

(a)基礎特徴間ゲイン調整型 (b)ボトムアップ-トップダウン間結合型

トップダウン要素による分類

(1)探索目標の事前学習型(2)文脈依存型(視覚探索以外のタスク,状態/状況依存)

5 5 5 5 5



視覚探索タスク

運転タスク

トップダウン要素による分類

(1)探索目標の事前学習型(2)文脈依存型(視覚探索以外のタスク,状態/状況依存)

技術要素による分類

(a)基礎特徴間ゲイン調整型 (b)ボトムアップ-トップダウン間結合型

# ボトムアップ型とトップダウン型



#### 技術要素による分類 (A)ゲイン調整型(B)BU-TD結合型



トップダウン要素による分類

(1)探索目標の事前学習型(2)文脈依存型(視覚探索以外のタスク,状態/状況依存)

技術要素による分類

(a)基礎特徴間ゲイン調整型 (b)ボトムアップ-トップダウン間結合型





### SNRの導入 Navalpakkam+, 2006



BU-TD間結合

ゲイン調整

目標探索

文脈依存

### SNRの導入 Navalpakkam+, 2006



BU-TD間結合

ゲイン調整

目標探索

文脈依存











#### 顕著度+顔領域 Cerf+, 2007



 $S = \frac{1}{4} \left( N(\overline{I}) + N(\overline{C}) + N(\overline{O}) + N(\overline{F}) \right)$ 

ゲイン調整

#### bottom-up salinecy

目標探索

文脈依存





#### GBVS + Viola & Jones

BU-TD間結合

J.





#### 顕著度+顔領域 Cerf+, 2007



ゲイン調整

目標探索

文脈依存

BU-TD間結合

J.

#### **Contextual Guidance** Torralba+, 2006



 $p(O,X \mid L,G) = \frac{1}{p(L \mid G)} p(L \mid O,X,G) p(X \mid O,G) p(O \mid G)$ 

O: 目標刺激のクラスラベル, X: 目標刺激の位置, L: 局所特徴, G: 大域特徴

#### **Contextual Guidance** Torralba+, 2006



 $p(O, X | L, G) = \frac{1}{p(L | G)} p(L | O, X, G) p(X | O, G) p(O | G)$ 

O: 目標刺激のクラスラベル, X: 目標刺激の位置, L: 局所特徴, G: 大域特徴

#### SUN (Saliency using natural statistics) Zhang+, 2008



$$p(O \mid L, X) = \frac{1}{p(L)} p(L \mid O) p(O \mid X)$$

$$\log p(O \mid L, X) \approx -\log p(L) + \log p(L \mid O)$$
$$= \log \frac{p(L,O)}{p(L)p(O)} = \log p(O \mid L) - \log p(O)$$

100

50









Model



People Absent



Model

Target: mugs



Bottom-up saliency



Contextual guidance



Subject consistency



Contextual modulation



SUN

#### SUN (Saliency using natural statistics) Zhang+, 2008





#### 探索非対称性(search asymmetries)



親和性が高い妨害刺激の中から新規性が高い目標刺激を 探索するのに要する時間は、その逆の場合より… 小さい

#### 探索非対称性(search asymmetries)



当てはまらない??ケースも多々ある

#### 0 \* BU + 1 \* TD



Human



#### 1 \* BU + 0 \* TD



0.5 \* BU + 0.5 \* TD



目標探索

文脈依存

ゲイン調整

BU-TD間結合

### 重みの評価 Zelinsky+, 2006







### 協調行動,共同注意 Ozeki+,2011



ゲイン調整

T

目標探索

文脈依存

BU-TD間結合

### 協調行動,共同注意 Ozeki+,2011



ゲイン調整

目標探索

文脈依存

BU-TD間結合










## 計算モデルの評価

## 計算モデルの評価尺度

Normalized scanpath saliency (NSS) ex. 注視点の顕著度と画像内の平均顕著度の差

Kullbach-Liebler (KL) distance

ex. 複数人注視点の顕著度分布v.s.ランダムサンプリングされた顕著度分布

#### ◆ 注視点分布と顕著性マップの相関

ex. 注視点分布=複数人の注視点分布へのガウシアンの畳み込み

#### ◆高顕著度領域への注視滞留頻度

ex. 高顕著度領域を抽出するための閾値と注視滞留頻度との関係を評価

#### 生成された注視点の目標捕捉率

ex. 注視点の移動回数と目標刺激の捕捉率との関係を評価

## モデル評価のためのデータセット

#### CRCNS eye-1

- 映像 :50シーン,640×480,60.27Hz,0:06-1:30, 風景,TVニュース,スポーツ,CM,ビデオゲーム タスク :主役への注意(事後タスクとしてインタビュー)
- 被験者:8名,23-32歳,女性3名,男性5名

視線 : ISCAN RK-464 eye-tracker, 240Hz





## モデル評価のためのデータセット LabelMe(open image labeling tool)

Objects: attic building car

## ソースコード

iLab Neuromorphic Vision C++ Toolkit

Graph-based visual saliency (GBVS)

Esaliency

Spectral residual

Frequency-tuned salient region detection

Saliency Toolbox

The bottom-up visual saliency of Itti+ to run on the Nokia N810 internet tablet

▶ Saliency mapをOpenCVで実装する







# 心理物理学的な特徴分析に関するReview Wolfe, 2004

Undoubted attributes		Probable attributes		
- Color - Motion - Orientation - Size		<ul> <li>Flicker</li> <li>Luminance pol</li> <li>Vernier offset</li> <li>Stereo depth &amp;</li> <li>Pictorial depth</li> </ul>	- Shape larity - Line termination - Closure & tilt - Topological status - cues - Curvature	
Possible attributes	D	oubtful cases	Probable non-attributes	
- Shading - Glossiness - Expansion - Number - Aspect ratio		<ul> <li>Novelty</li> <li>Letter identity</li> <li>Alphanumeric category</li> </ul>	<ul> <li>Intersection</li> <li>Optic flow</li> <li>Color change</li> <li>3D volumes</li> <li>Faces</li> <li>Your name</li> <li>Semantic category</li> </ul>	











本発表は、予稿ページ数や発表時間について 過分なご配慮を賜りました。 パターン認識・メディア理解研究会(PRMU) ヒューマン情報処理研究会(HIP)の 幹事の皆様に感謝致します。

本発表について、ご助言いただきました 柴田智広先生(奈良先端大)に感謝致します.