Incremental Sparse Saliency Detection

Yin Li, Yue Zhou, Lei Xu, Xiaochao Yang, Jie Yang
Institute of Image Processing & Pattern Recognition
Shanghai Jiao Tong University, China
Outline

- Introduction
- Related Work
- Our Proposed Method
- Experiments and Analysis
- Conclusion and Future Work
Introduction

Motivation

Everyone knows what attention is... — William James

- A computational approach to visual attention
- Fast selection for objects of interest in scenes
Introduction

Difficulties

- “Black box” problem
  - Covert & overt attention
  - Biological plausible

- Difficulty in evaluation
  - Quantitative analysis
  - The data set
Introduction

Overview

Original Image
Saliency Map
Proto Objects
Related Work

Related Work

Bruce2004, Hou2008,
Related Work

- **Other Method:**
  - Spectral Residual [Hou2007]
  - Contextual Guidance [Oliva2006]
  - Learning to Detect A Salient Object [Liu2007]

- ...
Our Proposed Model: Theory

_center-Surround Architecture_

Image

Video

Feature

\[ x = Fc \in \mathbb{R}^n \]

\[ S = [Fs_1, Fs_2, \ldots, Fs_N] \in \mathbb{R}^{n \times N} \]

*Overlapping allowed*
Our Proposed Method: Theory

- **Saliency as Incremental Coding Length (ICL)**
- **For certain lossy coding scheme** $L_\varepsilon(\cdot)$
  - $\varepsilon$ — distortion tolerance
- **Saliency of the center is defined as ICL:**
  \[
  \delta L_\varepsilon(x) = L_\varepsilon(S \cup x) - L_\varepsilon(S) = L_\varepsilon(x \mid S)
  \]
  \[
  Sa(x) = \delta L_\varepsilon(x)
  \]
- $x \mid S$ — encode $x$ with $S$
- **Optimum** coding scheme required
Our Proposed Method: Theory

Core Idea:

Saliency = Non-redundancy = Hard to encode
Our Proposed Method

Sparse Coding Scheme

- Center as the **sparse** linear representation of its surroundings

\[ x = \sum_{i=1}^{N} w_i F s_i = S w \quad w \in \mathbb{R}^N \]

Traditional approach

\[ w = \min_w \| x - S w \|_2^2 \]
Our Proposed Method

Sparse Coding Scheme

Our approach

\[ \min \| w \|_0 \quad s.t. \quad \| x - Sw \|_2^2 \leq \varepsilon \]

Optimum coding length under distortion \( \varepsilon \)

Computational intractable — NP hard
Our Proposed Method

Sparse Coding Scheme

- Our approach (NP-hard)

\[
\min \| w \|_0 \quad s.t. \quad \| x - Sw \|_2^2 \leq \varepsilon
\]

- Sparse assumption

\[
\| w \|_0 \ll N \quad \text{given} \quad n \ll N
\]

Solution (Polynomial)

\[
\min \| w \|_1 \quad s.t. \quad \| x - Sw \|_2^2 \leq \varepsilon
\]

*Feature invariance

(F is not important)
Our Proposed Method

Sparse Coding Scheme

Our solution

\[
\min \| w \|_1 \quad s.t. \quad \| x - Sw \|_2^2 \leq \varepsilon
\]

\[
\min \lambda \| w \|_1 + \frac{1}{2} \| x - Sw \|_2^2 \quad \lambda > 0
\]

Final saliency map by coding length

\[
Sa(c) = \delta L_\varepsilon (c) = \| w \|_0
\]
Our Proposed Method

Sparse Coding Scheme
Our Proposed Method: Summary

Summary

\begin{algorithm}
\textbf{Algorithm 1} (Incremental Sparse Saliency)

1. \textit{Input} : given image $I$
2. \textbf{for} each patch $c$ of the image $I$, calculate $x = Fc$
   and take patches from its surroundings to form $S$
   \begin{itemize}
   \item solve the optimization problem
     \[ \min \lambda \|w\|_1 + \frac{1}{2} \|x - Sw\|_2^2 \]
   \item given the sparse solution $w$, calculate the patch saliency $Sa(c)$ by $Sa(c) = \|w\|_0$, and accumulate the saliency by pixels
   \end{itemize}
3. \textit{end}
4. \textit{Output} : the saliency map of $I$
\end{algorithm}
One parameter: $\lambda > 0$

| Image | Human | $\lambda = 0.1$ | $\lambda = 0.2$ | $\lambda = 0.3$ | $\lambda = 0.4$ |
Experiment and Analysis: Images

From left to right
- Image
- Hand labeled
- Itti1998
- Hou2007
- Our Method
Experiment and Analysis: Video

Video

Saliency Map

Video

Saliency Map
Conclusion

- A visual saliency model by sparse coding
- Feature invariance
- Fairly good results

Future Work

- Quantitative evaluation of visual saliency
- Application of visual saliency in scene understanding
Thanks for your attention!