Locating Salient Group-Structured Image Features via Adaptive Compressive Sampling

Xingguo Li

Department of Electrical and Computer Engineering
University of Minnesota
Advisor: Prof. Jarvis Haupt

GlobalSIP
December 14, 2015

UNIVERSITY OF MINNESOTA

Background and Motivation

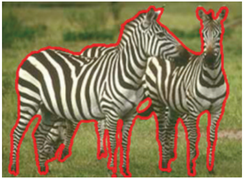
salient feature detection/localization in images

Broad applications in image processing, computer vision, surveillance etc.

salient feature detection/localization in images

Broad applications in image processing, computer vision, surveillance etc.

- foreground segmentation

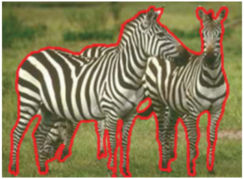


(AI & CV Lab., Seoul National University)

salient feature detection/localization in images

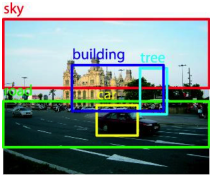
Broad applications in image processing, computer vision, surveillance etc.

- foreground segmentation



(AI & CV Lab., Seoul National University)

- object detection/recognition

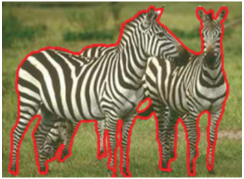


(PA & CV Dept., Italian Institute of Technology)

salient feature detection/localization in images

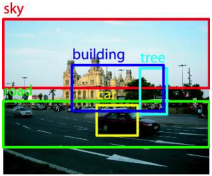
Broad applications in image processing, computer vision, surveillance etc.

- foreground segmentation



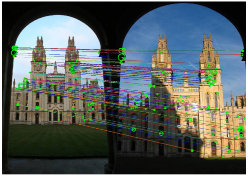
(AI & CV Lab., Seoul National University)

- object detection/recognition



(PA & CV Dept., Italian Institute of Technology)

- image matching

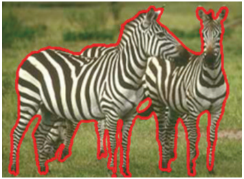


(Oxford Visual Geometry Group)

salient feature detection/localization in images

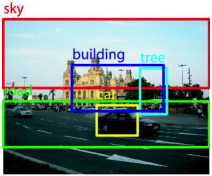
Broad applications in image processing, computer vision, surveillance etc.

- foreground segmentation



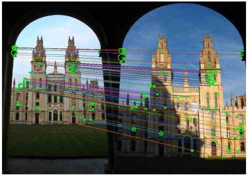
(AI & CV Lab., Seoul National University)

- object detection/recognition



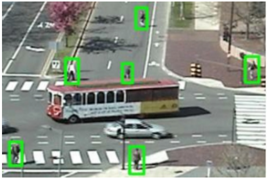
(PA & CV Dept., Italian Institute of Technology)

- image matching



(Oxford Visual Geometry Group)

- visual surveillance

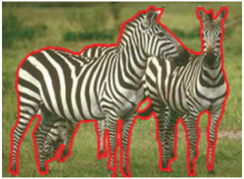


(Multimedia Lab, Chinese University of Hong Kong)

salient feature detection/localization in images

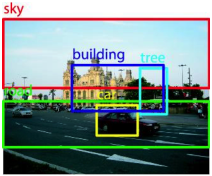
Broad applications in image processing, computer vision, surveillance etc.

- foreground segmentation



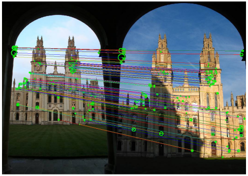
(AI & CV Lab., Seoul National University)

- object detection/recognition



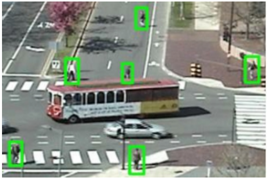
(PA & CV Dept., Italian Institute of Technology)

- image matching



(Oxford Visual Geometry Group)

- visual surveillance



(Multimedia Lab, Chinese University of Hong Kong)

- many more...

prior works

Bottom-up method: data-driven

- Contrast based: local contrast, global contrast (Itti et al. 1998, Achanta et al. 2009)
- Prior based: shape, location, background prior (Xie et al. 2013, Yang et al. 2013)
- Compressive Sensing based: low-rank homogeneous background + sparse salient foreground (Lang et al. 2012, Shen et al. 2013)

⋮

Top-down method: task dependent / goal driven

- Supervised learning (Liu et al. 2007)
- Dictionary learning (Yang et al. 2012)

⋮

prior works

Bottom-up method: data-driven

- Contrast based: local contrast, global contrast (Itti et al. 1998, Achanta et al. 2009)
- Prior based: shape, location, background prior (Xie et al. 2013, Yang et al. 2013)
- Compressive Sensing based: low-rank homogeneous background + sparse salient foreground (Lang et al. 2012, Shen et al. 2013)

⋮

Top-down method: task dependent / goal driven

- Supervised learning (Liu et al. 2007)
- Dictionary learning (Yang et al. 2012)

⋮

Drawback ! **FULL** imaging is required for feature/prior info. extraction.
Can be prohibitive in some applications, e.g., gigapixel photos.

our prior effort (Li & Haupt, IEEE Trans. Sig. Proc. 63(7) pp. 1792-1807, April 2015)

Key idea: locate salient features w/o fully imaging/reconstructing (bottom-up)

our prior effort (Li & Haupt, IEEE Trans. Sig. Proc. 63(7) pp. 1792-1807, April 2015)

Key idea: locate salient features w/o fully imaging/reconstructing (bottom-up)

- Raw image features (RGB or intensity): treat salient features as *outliers*.

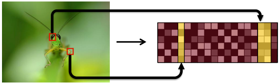


image patches $\xrightarrow{\text{vectorized}}$ columns of $\mathbf{M}$

our prior effort (Li & Haupt, IEEE Trans. Sig. Proc. 63(7) pp. 1792-1807, April 2015)

Key idea: locate salient features w/o fully imaging/reconstructing (bottom-up)

- Raw image features (RGB or intensity): treat salient features as *outliers*.

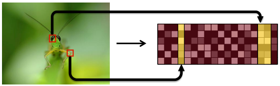
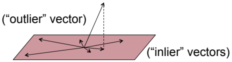


image patches $\xrightarrow{\text{vectorized}}$ columns of $\mathbf{M}$

- A two-step approach: assume matrices $\mathbf{M} \in \mathbb{R}^{n_1 \times n_2}$ admit a decomposition

$$\mathbf{M} = \underbrace{\mathbf{L}}_{\text{rank } r} + \underbrace{\mathbf{C}}_{k\text{-column sparse}}$$



our prior effort (Li & Haupt, IEEE Trans. Sig. Proc. 63(7) pp. 1792-1807, April 2015)

Key idea: locate salient features w/o fully imaging/reconstructing (bottom-up)

- Raw image features (RGB or intensity): treat salient features as *outliers*.

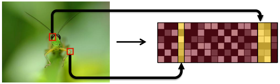
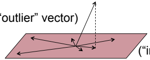


image patches $\xrightarrow{\text{vectorized}}$ columns of $\mathbf{M}$

- A two-step approach: assume matrices $\mathbf{M} \in \mathbb{R}^{n_1 \times n_2}$ admit a decomposition

$$\mathbf{M} = \underbrace{\mathbf{L}}_{\text{rank } r} + \underbrace{\mathbf{C}}_{k\text{-column sparse}}$$



Step 1 – *dimension reduction*: $\mathbf{Y}_{(1)} = \Phi\mathbf{MS} \quad (m \times \gamma n_2)$

convex demixing: $\text{argmin}_{\boldsymbol{L},\boldsymbol{C}} \ \|\boldsymbol{L}\|_* + \lambda\|\boldsymbol{C}\|_{1,2}$ s.t. $\mathbf{Y}_{(1)} = \boldsymbol{L} + \boldsymbol{C}$

our prior effort (Li & Haupt, IEEE Trans. Sig. Proc. 63(7) pp. 1792-1807, April 2015)

Key idea: locate salient features w/o fully imaging/reconstructing (bottom-up)

- Raw image features (RGB or intensity): treat salient features as *outliers*.

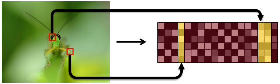
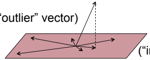


image patches $\xrightarrow{\text{vectorized}}$ columns of $\mathbf{M}$

- A two-step approach: assume matrices $\mathbf{M} \in \mathbb{R}^{n_1 \times n_2}$ admit a decomposition

$$\mathbf{M} = \underbrace{\mathbf{L}}_{\text{rank } r} + \underbrace{\mathbf{C}}_{k\text{-column sparse}}$$



Step 1 – *dimension reduction*: $\mathbf{Y}_{(1)} = \Phi\mathbf{M}\mathbf{S} \quad (m \times \gamma n_2)$

convex demixing: $\text{argmin}_{\boldsymbol{L},\boldsymbol{C}} \; \|\boldsymbol{L}\|_* + \lambda\|\boldsymbol{C}\|_{1,2}$ s.t. $\mathbf{Y}_{(1)} = \boldsymbol{L} + \boldsymbol{C}$

Step 2 – *orthogonal projection*: $\mathbf{y}_{(2)} = \phi \, \mathbf{P}_{\widehat{\mathcal{L}}^{\perp}_{(1)}} \Phi\mathbf{M}\mathbf{A}^T \; (1 \times p)$

sparse inference: solve $\widehat{\mathbf{c}} = \text{argmin}_{\mathbf{c}} \|\mathbf{c}_j\|_1$ s.t. $\mathbf{y}_{(2)} = \mathbf{c}\mathbf{A}^T$

our prior effort (Li & Haupt, IEEE Trans. Sig. Proc. 63(7) pp. 1792-1807, April 2015)

Key idea: locate salient features w/o fully imaging/reconstructing (bottom-up)

- Raw image features (RGB or intensity): treat salient features as *outliers*.

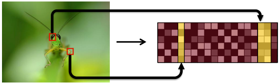
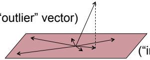


image patches $\xrightarrow{\text{vectorized}}$ columns of $\mathbf{M}$

- A two-step approach: assume matrices $\mathbf{M} \in \mathbb{R}^{n_1 \times n_2}$ admit a decomposition

$$\mathbf{M} = \underbrace{\mathbf{L}}_{\text{rank } r} + \underbrace{\mathbf{C}}_{k\text{-column sparse}}$$



Step 1 – *dimension reduction*: $\mathbf{Y}_{(1)} = \mathbf{\Phi}\mathbf{M}\mathbf{S} \quad (m \times \gamma n_2)$

convex demixing: $\text{argmin}_{\boldsymbol{L},\boldsymbol{C}} \ \|\boldsymbol{L}\|_* + \lambda\|\boldsymbol{C}\|_{1,2}$ s.t. $\mathbf{Y}_{(1)} = \boldsymbol{L} + \boldsymbol{C}$

Step 2 – *orthogonal projection*: $\mathbf{y}_{(2)} = \phi \ \mathbf{P}_{\widehat{\mathcal{L}}_{(1)}^{\perp}} \mathbf{\Phi}\mathbf{M}\mathbf{A}^T \ (1 \times p)$

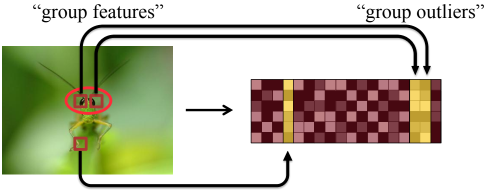
sparse inference: solve $\widehat{\mathbf{c}} = \text{argmin}_{\mathbf{c}} \|\mathbf{c}_j\|_1$ s.t. $\mathbf{y}_{(2)} = \mathbf{c}\mathbf{A}^T$

- Theoretical guarantee: $m\gamma n_2 + p = \mathcal{O}\left(r^2 \log r + k \log(n_2)\right)$ samples are sufficient for exact outlier identification w.h.p. (under structural assumptions)

Group Adaptive Compressive Sensing (GACS) for Salient Features

"group" features

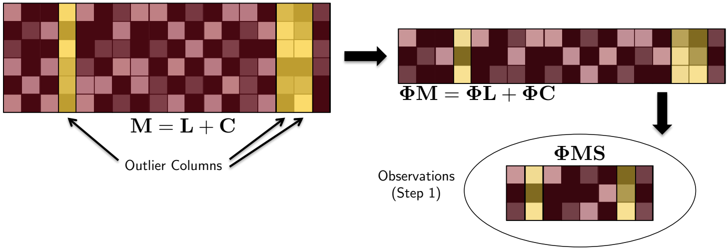
Salient features may be "grouped" in the pixel space



a two-step approach (step 1)

Collect Measurements: $\mathbf{Y}_{(1)} := \mathbf{\Phi M S} = \mathbf{\Phi}(\mathbf{L} + \mathbf{C})\mathbf{S}$ where

- $\mathbf{\Phi} \in \mathbb{R}^{m \times n_1}$ is a (random) measurement matrix ($m < n$)
- For $\gamma \in (0,1)$, $\mathbf{S}$ is a column sub matrix of identity with $\approx \gamma n_2$ columns *(rows sampled iid from a Bernoulli(γ) model)*

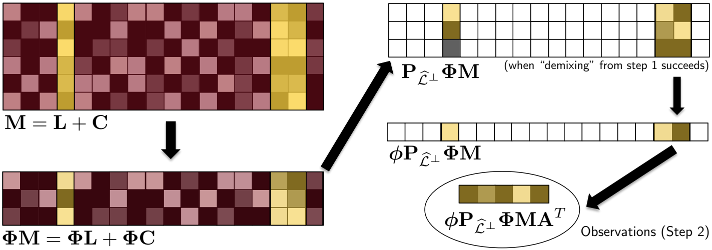


Apply *Outlier Pursuit* (Xu et al. 2012) to "pocket-sized" data $\mathbf{\Phi M S}$
(Idea: identify span of $\mathbf{\Phi L}$. Same 1st step as previous work)

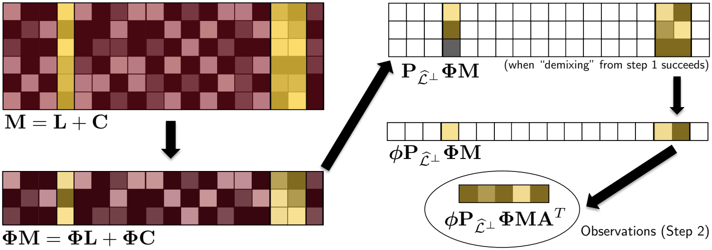
a two-step approach (step 2)

Collect measurements $\mathbf{y}_{(2)} := \phi\, \mathbf{P}_{\widehat{\mathcal{L}}^{\perp}_{(1)}} \mathbf{\Phi} \mathbf{M} \mathbf{A}^T$ where

- $\mathbf{\Phi} \in \mathbb{R}^{m \times n_1}$ is same (random) measurement matrix as in step 1,
- $\widehat{\mathcal{L}}_{(1)}$ is the linear subspace spanned by col's of $\widehat{\mathbf{L}}_{(1)}$ (learned in step 1)
- $\mathbf{P}_{\widehat{\mathcal{L}}_{(1)}}$ is orthogonal projector onto $\widehat{\mathcal{L}}_{(1)}$, and $\mathbf{P}_{\widehat{\mathcal{L}}^{\perp}_{(1)}} \triangleq \mathbf{I} - \mathbf{P}_{\widehat{\mathcal{L}}_{(1)}}$
- $\phi \in \mathbb{R}^{1 \times m}$ a random vector, $\mathbf{A} \in \mathbb{R}^{p \times n_2}$ a random matrix



a two-step approach (step 2)



Solve $\widehat{\mathbf{c}} = \operatorname{argmin}_{\mathbf{c}} \sum_{j=1}^{J} \|\mathbf{c}_j\|_2$ s.t. $\mathbf{y}_{(2)} = \mathbf{c}\mathbf{A}^T$

- group sparsity extension of previous work
- $\sum_{j=1}^{J} \|\mathbf{c}_j\|_2$ is a group norm
- J is the number of groups
- $\mathbf{c}_j \in \mathbb{R}^B$ is a subvector of $\mathbf{c} \in \mathbb{R}^{n_2}$, with $B = n_2/J$ as the size of each group
- $\mathrm{support}(\widehat{\mathbf{c}}) \triangleq \{i : \widehat{c}_i \neq 0\}$ is the estimate for outlier locations

Performance Analysis

structural "identifiability" assumptions

Def'n: (Column Incoherence Property)

Matrix $\mathbf{L} \in \mathbb{R}^{n_1 \times n_2}$ with $n_{\mathbf{L}} \leq n_2$ nonzero columns, rank r, and compact SVD $\mathbf{L} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ is said to satisfy the *column incoherence property* with parameter $\mu_{\mathbf{L}}$ if

$$\max_i \|\mathbf{V}^* \mathbf{e}_i\|_2^2 \leq \mu_{\mathbf{L}} \frac{r}{n_{\mathbf{L}}},$$

where $\{\mathbf{e}_i\}$ are basis vectors of the canonical basis for $\mathbb{R}^{n_2}$.

(small $\mu_{\mathbf{L}}$ precludes subspaces $\mathcal{L}$ defined by single col's of $\mathbf{L}$; an assumption that guarantees identifiability of $\{\mathbf{L}, \mathbf{C}\}$)

structural “identifiability” assumptions

Def’n: (Column Incoherence Property)

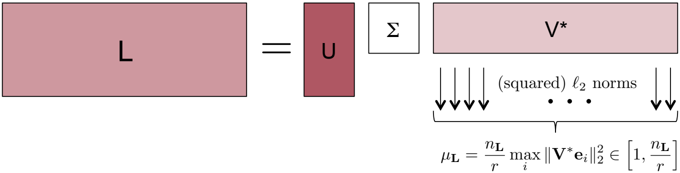
Matrix $\mathbf{L} \in \mathbb{R}^{n_1 \times n_2}$ with $n_{\mathbf{L}} \leq n_2$ nonzero columns, rank r, and compact SVD $\mathbf{L} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^*$ is said to satisfy the *column incoherence property* with parameter $\mu_{\mathbf{L}}$ if

$$\max_i \|\mathbf{V}^* \mathbf{e}_i\|_2^2 \leq \mu_{\mathbf{L}} \frac{r}{n_{\mathbf{L}}},$$

where $\{\mathbf{e}_i\}$ are basis vectors of the canonical basis for $\mathbb{R}^{n_2}$.

(small $\mu_{\mathbf{L}}$ precludes subspaces $\mathcal{L}$ defined by single col’s of $\mathbf{L}$; an assumption that guarantees identifiability of $\{\mathbf{L}, \mathbf{C}\}$)

Graphically:



provable recovery

Structural conditions: **(Xu et al. 2012)**

Suppose components $\mathbf{L}$ and $\mathbf{C}$ satisfy the structural conditions: (1) $\mathrm{rank}(\mathbf{L}) = r$, (2) $\mathbf{L}$ has $n_{\mathbf{L}} \leq n_2$ nonzero columns, (3) $\mathbf{L}$ satisfies the *column incoherence property* with parameter $\mu_{\mathbf{L}}$, and (4) $|\mathcal{I}_{\mathbf{C}}| = k$.

Theorem: **(Li & Haupt, GlobalSIP, 2015)**

For any $\delta \in (0,1)$, take

$$k \leq n_2/(c_1 r \mu_{\mathbf{L}}), \quad \gamma \geq c_2 r \mu_{\mathbf{L}} \log r / n_{\mathbf{L}},$$

$$m \geq c_3(r + \log k), \quad p \geq c_4 \left(k + (k/\sqrt{B}) \log((n_2 - k)/B)\right).$$

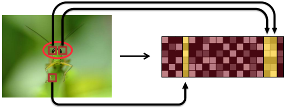
let ϕ have elements drawn iid from any continuous distribution, and take the outlier pursuit reg. parameter $\lambda = \frac{3}{7\sqrt{k_{\mathrm{ub}}}}$, where k_{ub} is any upper bound of k. The following hold simultaneously w.p. $\geq 1 - 3\delta$: the support estimate produced by our method is correct, and the no. of obs. is no greater than

$$\underbrace{(3/2)\gamma m n_2 + p}_{\text{as few as } \mathcal{O}((r+\log k)(\mu_{\mathbf{L}} r \log r) + k + \frac{k}{\sqrt{B}} \log \frac{n_2}{B})}$$

Experimental Results

grouping effect

Recall: vectorize (non-overlap) image patches into columns of **M**

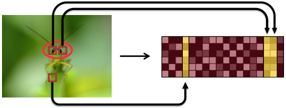


Advantage of grouping features: lower sample demands

$$\mathcal{O}((r + \log k)(\mu_{\mathbf{L}} r \log r) + k + \frac{k}{\sqrt{B}} \log \frac{n_2 - k}{B}) \text{ vs.}$$
$$\mathcal{O}((r + \log k)(\mu_{\mathbf{L}} r \log r) + k \log \frac{n_2}{k})$$

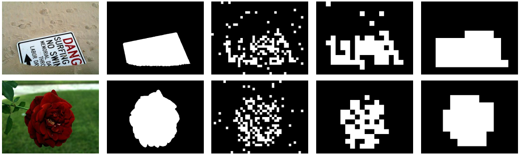
grouping effect

Recall: vectorize (non-overlap) image patches into columns of **M**



Advantage of grouping features: lower sample demands

$$\mathcal{O}((r + \log k)(\mu_{\mathbf{L}} r \log r) + k + \frac{k}{\sqrt{B}} \log \frac{n_2 - k}{B}) \text{ vs.}$$
$$\mathcal{O}((r + \log k)(\mu_{\mathbf{L}} r \log r) + k \log \frac{n_2}{k})$$



(a) (b) (c) (d) (e)

Detection results with the grouping effect. (a) original images; (b) ground truth; detection result (c) w/o grouping ($B = 1$) and with grouping effects; (d) $B = 2$; and (e) $B = 3$. Sampling rate: 2.5% ($\gamma = 0.2$, $m = 0.1 n_1$ and $p = 0.5 n_2$).

low-level image features

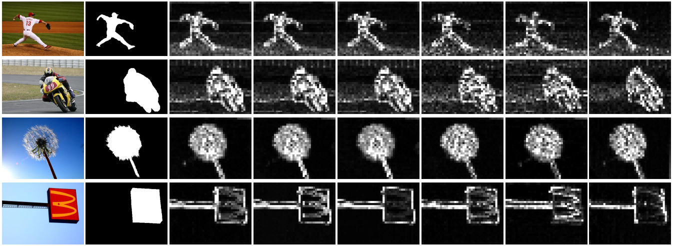
Each step of our two-step process obtains linear measurements of the image pixels.

⇒ Can incorporate any linear "preprocessing" (e.g., *filtering*) into the overall measurement model at the feature acquisition stage.

low-level image features

Each step of our two-step process obtains linear measurements of the image pixels.

⇒ Can incorporate any linear "preprocessing" (e.g., *filtering*) into the overall measurement model at the feature acquisition stage.



(a) (b) (c) (d) (e) (f) (g) (h)

Gray scale (maginitude of entries of $\hat{\mathbf{c}}$) saliency map estimation.(a) original images; (b) ground truth; (c)-(e) RGB planes individually; filtered intensity images with (f) Laplacian of Gaussian filter, (g) Horizontal Edge filter and (h) Vertical Edge filter. Sampling rate: 4.5% ($\gamma = 0.2$, $m = 0.2n_1$, $p = 0.5n_2$, $n_1 = 100$ and $n_2 = 1200$)

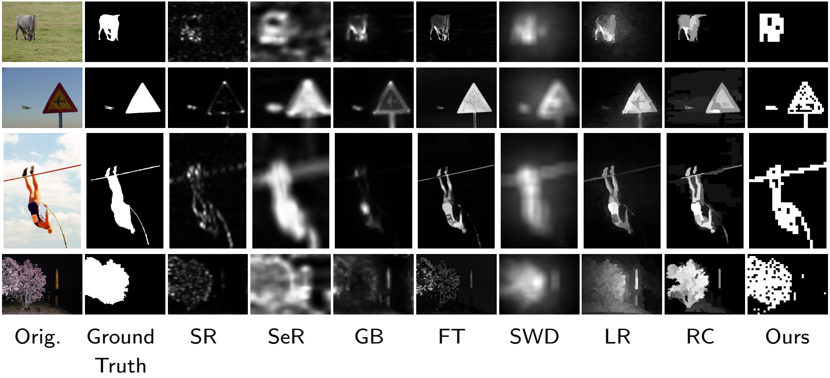
comparisons w/existing saliency detection methods

State-of-the-art methods:

- spectral residual (SR) (Hou & Zhang 2007)
- self-resemblance (SeR) (Seo & Milanfar 2009)
- global based (GB) (Harel et al. 2006)
- frequency tuned (FT) (Achanta et al. 2009)
- spatially weighted dissimilarity (SWD) (Duan et al. 2011)
- low rank (LR) (Shen & Wu 2012)
- region contract (RC) (Cheng et al. 2014)

Database: MSRA10K (Cheng et al. 2014)

comparisons w/existing saliency detection methods

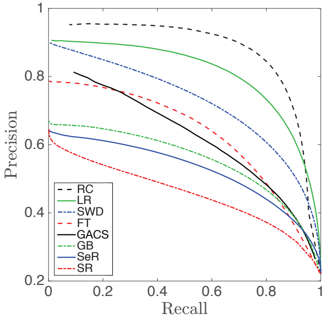


Detection results for the MSRA10K Salient Object Database for various methods. For our approach, the results correspond to G, LoG, I, and R respectively from top to bottom. Sampling rate: 2.5% on average.

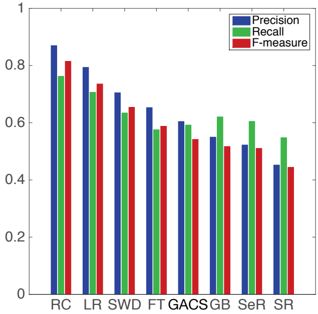
comparisons w/existing saliency detection methods

More results:

- Precision: $P = \frac{TP}{TP+FP}$, TP: true positive, FP: false positive
- Recall: $R = \frac{TP}{TP+FN}$, FN: false negative
- F-measure $= \max_{P,R} \frac{(\beta^2+1)P \cdot R}{(\beta^2 P+R)}$, $\beta^2 = 0.3$



(a) Precision-Recall curve



(b) F-measure

Conclusions

final comments

Direct saliency localization is possible (w/o full imaging)

- Low sample complexity
- Low computational complexity

Extensions under examination:

- Non-linear "post-processing" of image features
- Observation with missing data

Current investigation:

- Seek known patterns embedded in unknown backgrounds (Where's Waldo?)
- Stability analyses (e.g., in noisy settings or when data are missing or both)

Techniques like GACS may become increasingly **IMPORTANT** when data becomes bigger and bigger!

thanks!

Advisor/Coauthor: Prof. Jarvis Haupt

Research Support:

NSF Award No. CCF-1217751 *(Exploiting Saliency in Compressive and Adaptive Sensing)*

Thanks!

lixx1661@umn.edu
http://www.tc.umn.edu/~lixx1661/

Extra Slides

non-linear “post-processing”

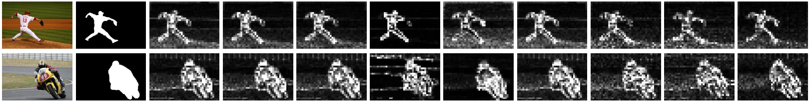
Further exploration of feature extraction, e.g., “stacked HSI”
(RGB to HSI on the compressed data $\mathbf{\Phi M}$)

Overall procedure of feature acquisition, up to $\mathbf{\Phi M}$, is still linear

non-linear "post-processing"

Further exploration of feature extraction, e.g., "stacked HSI"
(RGB to HSI on the compressed data $\mathbf{\Phi M}$)

Overall procedure of feature acquisition, up to $\mathbf{\Phi M}$, is still linear



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k)

Gray scale (maginitude of entries of $\hat{\mathbf{c}}$) saliency map estimation. (a) original images; (b) ground truth; (c)-(e) RGB planes; (f)-(h) stacked HSI individually; filtered intensity images with (i) Laplacian of Gaussian filter, (j) Horizontal Edge filter and (k) Vertical Edge filter. Sampling rate: 4.5% ($\gamma = 0.2$, $m = 0.2n_1$, $p = 0.5n_2$, $n_1 = 100$ and $n_2 = 1200$)