Implicit Rank-Sparsity Decomposition: Applications to Saliency/Co-Saliency Detection

Yi-Lei Chen
Depart of Computer Science
National Tsing Hua University
Hsinchu, Taiwan
fallcolor@gmail.com

Chiou-Ting Hsu*
Depart of Computer Science
National Tsing Hua University
Hsinchu, Taiwan
cthsu@cs.nthu.edu.tw

***Abstract*—Modern techniques rely on convex relaxation to derive tractable approximations for rank-sparsity decomposition. However, the resultant precision loss usually deteriorates the performance in real-world applications. In this paper, we focus on the topic of visual saliency detection and consider the inherent uncertainty existing in observations, which may originate from both low-rank and sparse components. We formulate the rank-sparsity model with an implicit weighting factor and show that this weighting factor characterizes the nature of visual saliency. The proposed model is generalized to solve saliency and co-saliency detection in a unified way. In addition, this model can easily incorporate center-prior or other top-down priors and can extend to multi-task learning to explore the interrelation between multiple features. Experimental results demonstrate that our method improves existing rank-sparsity decomposition, and also outperforms most state of the arts on two salient object databases.**

Keywords—saliency detection; co-saliency detection; rank-sparsity decomposition

I. INTRODUCTION (*HEADING 1*)

Over the last half decade, the increasing popularity of low-rank matrix recovery shows its significance and provides a theoretic foundation for a variety of computer vision tasks. Through recent developments, one of the most important technical extensions is the rank-sparsity decomposition. With the so-called rank-sparsity incoherence [1], one can use convex optimization to decompose a corrupted matrix into two components: a clean matrix exhibiting low-rank structure and a sparse matrix containing gross errors. This decomposition scheme inherently characterizes ubiquitous outliers in real-world data, and thus has been successfully applied to image batch alignment [2], key-point detection [3], photometric stereo [4], saliency detection [5-7], and so on. Although convex relaxation has been popularly adopted to make this decomposition problem tractable, it would also lose the problem precision. To improve the estimate on rank-sparisty decomposition, some research [8] focused on non-convex formulation; whereas others [9-10] incorporated application-dependent regularizers, which are usually non-convex but did improve their performance.

Instead of using relaxation technique in decomposition scheme, we aim to balance the tradeoff between problem precision and model feasibility so as to tackle real-world applications. In this paper, we focus on the topic of visual saliency detection. Visual saliency captures the attention of human observers and its detection plays a significant role for many subsequent vision tasks (e.g., object detection, image editing). To date, much research has pointed out that computational models can be used to detect image saliency via low-level visual cues [11-12]. Existing methods can be classified into three categories, which utilize *contrast*, *uniqueness*, and *frequency analysis* to measure image saliency. Contrast-based methods assumed that image saliency reveals high contrast different from their surroundings. Different definitions of surround thus relate to local contrast [13-14] or global contrast [15-17] based strategies. On the other hand, frequency-based methods [18-21] detect image saliency via suppressing non-saliency in the log spectrum. The category of uniqueness-based methods is the most diverse one but still share similar perspective: saliency shows unique and different behavior in comparison with non-saliency. Consequently, self-information [22], graph model [23], incremental coding length [24], and patch rarity [25] have been utilized to capture these unique pixels. In addition, with the recent advance in rank-sparsity decomposition, [5-7] proposed to detect the uniqueness corresponding to sparse errors in the decomposition model. However, because of the convex relaxation technique, their methods lost the precision from relaxing *sparseness* (l_0-norm) to *sparse component* (l_1-norm) and thus can not explicitly determine the saliency score.

Rather than working on only the low-level saliency, some literature integrated top-down information from learning or prior knowledge [26-27] to model human perception. For example, the co-saliency detection scenario assumes that common image saliency exists in both two given images and should better reveal salient information than their individually detected saliency. However, to the best of our knowledge, only little research [28-29] investigated this issue. A unified solution is needed to establish the link from saliency to co-saliency detection. In this paper, we first consider the uncertainty of observation to reduce the precision loss in existing rank-sparsity decomposition. By introducing an implicit weighting factor, we show that the new model characterizes the nature of visual saliency, and solves saliency and co-saliency detection in a unified way. The proposed model can effortlessly incorporate center-prior or other top-down priors and extend to multi-task learning to explore the interrelation between multiple features. In the experiments, our method outperforms most state of the arts (including the two methods [6-7] which are also based on rank-sparsity decomposition). Furthermore,

1051-4651/14 $31.00 © 2014 IEEE
DOI 10.1109/ICPR.2014.400

co-saliency information, center-prior, and multi-task learning all improve the detection accuracy under our model.

The rest of this paper is organized as follows. Section II briefly reviews relate work and detail our proposed rank-sparsity decomposition. Section III explains how we apply the proposed model to saliency and co-saliency detection. Section IV shows the experimental results and comparison with state-of-the-art methods. Finally, section V gives the conclusion.

II. Implict Rank-Sparsity Decomposition

A. Notations

We first summarize the notations used in this paper. Lower case letters $(x, y, ...)$ denote scalar, bold lower case letters $(\mathbf{x}, \mathbf{y}, ...)$ denote vector, and bold upper case letters $(\mathbf{X}, \mathbf{Y}, ...)$ denote matrix. x_i is the i^{th} element of $\mathbf{x}$, and $x_{i,j}$ and $\mathbf{x}_j$ are the $(i,j)^{\text{th}}$ element and the j^{th} column of $\mathbf{X}$, respectively. $\text{Diag}(\mathbf{x})$ denotes a diagonal matrix, and $\text{trace}(\mathbf{X})$ denotes the sum of all diagonal elements in $\mathbf{X}$. The singular value decomposition (SVD) of $\mathbf{X}$ is written by $\mathbf{U}_\mathbf{X}\mathbf{\Sigma}_\mathbf{X}\mathbf{V}_\mathbf{X}^\text{T}$.

The nuclear norm $\|\mathbf{X}\|_*$ is defined by $\sum_i \sigma_i(\mathbf{X})$, where $\sigma_i(\mathbf{X})$ is the i^{th} largest singular value of $\mathbf{X}$. The l_0-norm of $\mathbf{X}$ indicates the number of nonzero $x_{i,j}$, and $\|\mathbf{X}\|_1$, $\|\mathbf{X}\|_{2,1}$ are defined by $\sum_{i,j}|x_{i,j}|$ and $\sum_j \sqrt{\sum_i (x_{i,j})^2}$, respectively. Finally, $\|\mathbf{X}\|_\text{F}$ is defined by $\sqrt{\sum_{i,j}(x_{i,j})^2}$.

B. Related Work

Modern techniques of rank-sparsity decomposition can be traced back to Robust Principal Component Analysis (RPCA) [30], where the authors first proposed a convex relaxation to decompose a corrupted matrix $\mathbf{X}$ into a low-rank matrix $\mathbf{R}$ and a sparse matrix $\mathbf{E}$:

$$\hat{\mathbf{R}}, \hat{\mathbf{E}} = \text{argmin}\ \|\mathbf{R}\|_* + \lambda\|\mathbf{E}\|_1 \quad \text{s.t.}\ \mathbf{X} = \mathbf{R} + \mathbf{E}\,. \quad (1)$$

In Equation (1), the authors replaced $\text{rank}(\mathbf{R})$ and $\|\mathbf{E}\|_0$ by their tightest convex surrogates $\|\mathbf{R}\|_*$ and $\|\mathbf{E}\|_1$, respectively, to make the optimization tractable. Many variants of RPCA have also been developed to solve different vision/learning tasks. For example, the low-rank representation (LRR):

$$\hat{\mathbf{Z}}, \hat{\mathbf{E}} = \text{argmin}\ \|\mathbf{Z}\|_* + \lambda\|\mathbf{E}\|_{2,1} \quad \text{s.t.}\ \mathbf{X} = \mathbf{XZ} + \mathbf{E}\,, \quad (2)$$

which further considers both low-rank subspaces spanned by $\mathbf{X}$ and sample-specific errors captured by $l_{2,1}$-norm, was recently reported for saliency detection [6] and robust clustering [31].

Despite the encouraging results of RPCA, researchers still seek better solutions to tackle its two major limitations. First, the l_1-norm would bias the estimates of $\mathbf{E}$ when the corrupted entries are not sparse and random enough. Second, some applications are more concerned with the *sparseness* of $\mathbf{E}$ than its exact values. For example, in the foreground detection problem [10], because the goal is to locate the foreground but not to estimate the foreground itself, the authors proposed to incorporate a binary spatial support to measure the *sparseness* $\|e_{i,j}\|_0$ and also include spatial coherency as their sparseness priors. They model the rank-sparsity decomposition by Markov random field (MRF) and solve it by the graph-cut algorithm. Although [10] outperforms RPCA on synthetic data and foreground detection problem, their formulation may still lose the precision on *sparseness* because they utilizes discrete optimization to predict binary supports. From a more general perspective, the meaning of *sparseness* should be related to a continuous random variable rather than a discrete label. For example, in visual saliency detection, a continuous support between 0 and 1 should better explain a local observation with respect to its surroundings than a binary value.

C. Proposed Model

To reduce the precision loss in *sparseness* measure, we propose a novel implicit rank-sparsity decomposition. We first introduce a general form to clarify the relationship between our model and existing methods:

$$\begin{aligned}\mathbf{X} = \mathbf{R} + \mathbf{E} &= \mathbf{R} + \rho\mathbf{E} + (1-\rho)\mathbf{E} \\ &= \mathbf{R} + \rho\mathbf{E} + (1-\rho)(\mathbf{X} - \mathbf{R})\,. \end{aligned} \quad (3)$$

Theoretically, RPCA-based methods focus on the modeling of Equation (3) when $\rho = 1$, whereas [10] considers the case when ρ is either 0 or 1. In contrast, considering the uncertainty existing in real-world data, we propose to explicitly include this uncertainty by assuming $0 \le \rho \le 1$ in order to achieve more accurate decomposition.

To include the uncertainty into the rank-sparsity decomposition, we introduce a weighting factor $\mathbf{P}$, where each element $p_{i,j}$ in $\mathbf{P}$ indicate the uncertainty of observation $x_{i,j}$ originated from both low-rank and sparse components:

$$\begin{aligned}&\hat{\mathbf{R}}, \hat{\mathbf{P}} = \text{argmin}\ \alpha\|\mathbf{R}\|_* + \beta\textstyle\sum_{i,j} p_{i,j} + \\ &\gamma\textstyle\sum_{i,j,k,l}(p_{i,j} - p_{k,l})^2 w_{i,j,k,l} + \sum_{i,j}(1-p_{i,j})(x_{i,j} - r_{i,j})^2 \\ &\qquad\qquad \text{s.t.}\ 0 \le p_{i,j} \le 1\quad .\end{aligned} \quad (4)$$

In Equation (4), the 2nd term measures the weighted l_0-norm and the 3rd term encodes the pairwise relation defined according to the *a priori* knowledge about the sparseness. A Gaussian noise model $x_{i,j} = r_{i,j} + n_{i,j}$, which is defined by the 4th term, is also included to tackle the noise in real observation.

III. Applications in Visual Saliency Detection

Assuming non-salient regions usually share similar visual characteristics and salient regions have very different behaviors from non-salient regions, one can utilize low-rank and sparse properties to characterize non-salient and salient regions, respectively. Existing methods [5-7] rely on RPCA-based models to estimate the sparse component $\mathbf{E}$ and then determine the saliency score from $\mathbf{E}$. However, this two-stage processing may bias the estimate and thus loses the detection precision.

In this section, we demonstrate how to apply our implicit rank-sparsity model to explicitly estimate the saliency score. We will also show that, in addition to the sparseness priors defined in Equation (4), our model can easily incorporate other priors (e.g., center bias effect, co-saliency information) in a unified way and can also extend to multi-task learning.

A. Saliency Detection

Our goal is to explicitly estimate the saliency score in patch-level. We divide an image into N non-overlapping patches and represent each patch by a k–dimensional feature

vector. Let $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{k \times N}$ denote the feature matrix, $\mathbf{p} \in \mathbb{R}^N$ denote the patch-based support, $\mathbf{x}_i \in \mathbb{R}^k$ be the visual feature of the i^{th} patch, and p_i denotes its corresponding saliency score. We formulate the saliency detection problem as

$$\widehat{\mathbf{R}}, \widehat{\mathbf{p}} = \operatorname{argmin} \alpha\|\mathbf{R}\|_* + \beta \textstyle\sum_i p_i + \gamma \textstyle\sum_{i,j}(p_i - p_j)^2 w_{i,j} + \textstyle\sum_i (1 - p_i)\|\mathbf{x}_i - \mathbf{r}_i\|^2 \quad \text{s.t. } 0 \le p_i \le 1 \quad (5)$$

It is worth noting that Equation (5) inherently includes the column-wise (i.e., patch-wise) sparsity, which was additionally measured by $l_{2,1}$-norm in [6]. The pairwise weight $w_{i,j}$ controls the sparseness priors and is critical to the performance. Here, we consider both feature similarity and spatial coherency between two patches to define $w_{i,j}$:

$$w_{i,j} = \begin{cases} \widehat{w}_{i,j}, & \text{if } \mathbf{x}_i = \text{knn}(\mathbf{x}_j) \text{or } \mathbf{x}_j = \text{knn}(\mathbf{x}_i) \\ 0, & \text{otherwise} \end{cases}, \text{ and}$$

$$\widehat{w}_{i,j} = \delta \exp\left(\frac{-\|\mathbf{c}_i - \mathbf{c}_j\|^2}{\sigma_{\mathbf{c}}^2}\right) + (1 - \delta) \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma_{\mathbf{x}}^2}\right), \quad (6)$$

where $\mathbf{c}_i$ denotes the coordinate of centroid of the i^{th} patch, $\text{knn}(\cdot)$ indicates the k nearest neighbor function, and δ is used to balance the impacts between feature similarity and spatial coherency (k=8 and δ=0.5 in our experiments).

Although Equation (5) is non-convex and difficult to solve, if we solve either $\mathbf{R}$ or $\mathbf{p}$ with the other one fixed, the two corresponding sub-problems are convex and can be solved exactly. We thus solve $\mathbf{R}$ and $\mathbf{p}$ iteratively until the changes of $\mathbf{R}$ and $\mathbf{p}$ between consecutive iterations are smaller than a predefined threshold. Below we detail the optimization.

1) Optimization of $\boldsymbol{R}$

By ignoring those terms independent of $\mathbf{R}$, we have a unconstrained minimization problem:

$$\widehat{\mathbf{R}} = \operatorname{argmin} \alpha\|\mathbf{R}\|_* + \text{trace}((\mathbf{X} - \mathbf{R})\mathbf{\Lambda}(\mathbf{X} - \mathbf{R})^{\mathrm{T}}), \quad (7)$$

where $\mathbf{\Lambda} = \text{Diag}([1 - p_1, \dots, 1 - p_N]^{\mathrm{T}})$. To make Equation (7) tractable, we adopt the linearization technique:

$$\widehat{\mathbf{R}} = \operatorname{argmin} \alpha\|\mathbf{R}\|_* + \frac{\tau}{2}\left\|\mathbf{R} - (\mathbf{R}^t - \frac{1}{\tau}\nabla_{\mathbf{R}}(\mathbf{R}^t))\right\|_{\mathrm{F}}^2, \quad (8)$$

where $\mathbf{R}^t$ is the approximation obtained at the previous step, τ denotes $\sigma_1(\mathbf{\Lambda} + \mathbf{\Lambda}^{\mathrm{T}})$, and $\nabla_{\mathbf{R}}(\mathbf{R}^t) = 2(\mathbf{R}^t - \mathbf{X})\mathbf{\Lambda}$. Because $\mathbf{\Lambda}$ is diagonal and its singular values are equal to the diagonal elements, τ is equal to $2 - 2p_{\min}$ ($p_{\min}$ is the minimum element of $\mathbf{p}$). We thus rewrite Equation (8) and have

$$\widehat{\mathbf{R}} = \operatorname{argmin} \alpha\|\mathbf{R}\|_* + \frac{\tau}{2}\|\mathbf{R} - \mathbf{M}\|_{\mathrm{F}}^2, \text{ where}$$

$$\mathbf{M} = \mathbf{R}^t\left(\mathbf{I} - \widetilde{\mathbf{\Lambda}}\right) + \mathbf{X}\widetilde{\mathbf{\Lambda}}, \text{ and}$$

$$\widetilde{\mathbf{\Lambda}} = \mathbf{\Lambda}/(1 - p_{\min}), \mathbf{I} \text{ is identity matrix.} \quad (9)$$

From [32], the optimum of Equation (9) is derived by SVD shrinkage technique:

$$\widehat{\mathbf{R}} = \mathbf{U}_{\mathbf{M}}\widetilde{\mathbf{\Sigma}}_{\mathbf{M}}\mathbf{V}_{\mathbf{M}}^{\mathrm{T}}, \quad (10)$$

where $\widetilde{\mathbf{\Sigma}}_{\mathbf{M}}$ denotes the shrunk $\mathbf{\Sigma}_{\mathbf{M}}$ with all shrunk diagonal elements; i.e., $\max(\sigma_i(\mathbf{M}) - \alpha/\tau, 0)$.

2) Optimization of $\boldsymbol{p}$

By ignoring the terms independent of $\mathbf{p}$, we have a constrained minimization problem:

$$\widehat{\mathbf{p}} = \operatorname{argmin} \mathbf{y}^{\mathrm{T}}\mathbf{p} + \gamma\mathbf{p}^{\mathrm{T}}\mathbf{L}\mathbf{p} \quad \text{s.t. } 0 \le \mathbf{p} \le 1, \quad (11)$$

where $\mathbf{y} = [\beta - \|\mathbf{x}_1 - \mathbf{r}_1\|^2, \dots, \beta - \|\mathbf{x}_N - \mathbf{r}_N\|^2]^{\mathrm{T}}$, and $\mathbf{L}$ is the graph Laplacian matrix defined by $\mathbf{D} - \mathbf{W}$ ($\mathbf{D}$ denotes a diagonal matrix with $d_{i,i} = \sum_j w_{i,j}$). Equation (11) is a typical quadratic programming. In our experiments, we solve it using the Matlab toolbox *quadprog()*.

B. Model Extension

1) Center-prior and beyond

Top-down semantic information and center-prior have been used to facilitate saliency detection [26-27]. Here, we adopt the simple center-prior, also called *center bias effect*, to demonstrate how we incorporate other priors into our model. The *center bias effect* assumes that human attention usually concentrates on image center and thus would bias the detected saliency score. We define an off-center-bias support $\mathbf{q} \in \mathbb{R}^N$ by

$$q_i = 1 - \exp\left(\frac{-\|\mathbf{c}_i - \bar{\mathbf{c}}\|^2}{\sigma_{\mathbf{c}}^2}\right), \quad (12)$$

where $\bar{\mathbf{c}}$ is the spatial coordinate of image center. To include the center-bias prior, we replace the 2nd term in Equation (5) by $\beta \sum_i q_i p_i$. Then $\mathbf{y}$ in Equation (11) is accordingly modified as

$$\mathbf{y} = [\beta q_1 - \|\mathbf{x}_1 - \mathbf{r}_1\|^2, \dots, \beta q_N - \|\mathbf{x}_N - \mathbf{r}_N\|^2]^{\mathrm{T}}. \quad (13)$$

If q_i is small (i.e., the i^{th} patch locates close to image center), then its sparseness penalty is reduced and this patch is more likely to be a salient patch. Our proposed strategy is intuitive and simple. One only needs to change the definition of support $\mathbf{q}$ to incorporate different prior information.

2) Multi-task learning

When considering each type of feature used in saliency detection as an individual task, the multi-task learning could be used to further improve the detection performance [6]. Assume we have l feature types and represent the feature matrix by $\mathbf{X} = [\mathbf{X}_1^{\mathrm{T}}, \dots, \mathbf{X}_l^{\mathrm{T}}]^{\mathrm{T}}$. We extend our method to multi-task framework by relaxing the single low-rank model defined in Equation (7) into l low-rank models:

$$\widehat{\mathbf{R}_1}, \dots, \widehat{\mathbf{R}_l} = \operatorname{argmin} \textstyle\sum_{i=1}^{l}(\alpha\|\mathbf{R}_i\|_* + \text{trace}((\mathbf{X}_i - \mathbf{R}_i)\mathbf{\Lambda}(\mathbf{X}_i - \mathbf{R}_i)^{\mathrm{T}})). \quad (14)$$

In Equation (14), each $\mathbf{R}_i$ is independent of the others and can be solved individually. With this relaxation, different types of visual features can retain their different low-rank structures (which is usually more faithful to real cases) while still sharing consistent saliency scores.

3) Co-Saliency detection

We extend Equation (5) to tackle the co-saliency detection problem. Given two images, we first compute their feature matrices $\mathbf{X}^1 \in \mathbb{R}^{k \times N_1}$, $\mathbf{X}^2 \in \mathbb{R}^{k \times N_2}$ and estimate the two underlying low-rank matrices $\mathbf{R}^1, \mathbf{R}^2$ and also the concatenated support $\mathbf{p} \in \mathbb{R}^{N_1 + N_2}$ by

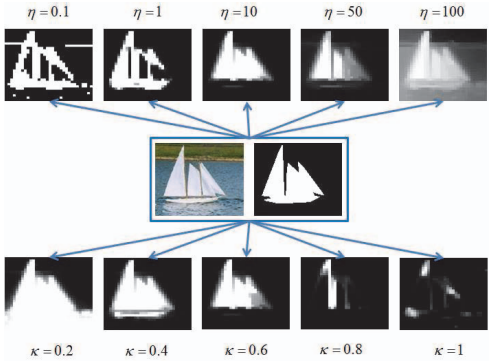


Figure 1: Our detection results using different κ (fixed $\eta = 10$ in the lower row) and η (fixed $\kappa = 0.5$ in the upper row).

$$\widehat{\mathbf{R}^1}, \widehat{\mathbf{R}^2}, \hat{\mathbf{p}} = \operatorname{argmin} \alpha \sum_{i=1,2} \left\| \mathbf{R}^i \right\|_* + \beta \sum_i p_i + \gamma \sum_{i,j} (p_i - p_j)^2 w_{i,j} + \sum_{i=1}^{N_1} (1 - p_i) \|\mathbf{x}_i^1 - \mathbf{r}_i^1\|^2 + \sum_{i=1}^{N_2} (1 - p_{i+N_1}) \|\mathbf{x}_i^2 - \mathbf{r}_i^2\|^2 \quad \text{s.t. } 0 \le p_i \le 1 \quad (15)$$

The major difference lies in *sparseness priors* term, where both intra- and inter-image patch similarity are considered for co-saliency detection. We rewrite the weight matrix $\mathbf{W}$ by:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}^{11} & \mathbf{W}^{12} \\ \mathbf{W}^{21} & \mathbf{W}^{22} \end{bmatrix}, \quad (16)$$

where $\mathbf{W}^{21}$ is the transpose of $\mathbf{W}^{12}$, and $\mathbf{W}^{11}$ and $\mathbf{W}^{22}$ are intra-image weight matrices (derived by Equation (6)) of the two input images, respectively. To define $\mathbf{W}^{12}$, we simply utilize the feature similarity because now we have no priors about their spatial correspondence:

$$w_{i,j}^{12} = \begin{cases} \exp\left(\frac{-\|\mathbf{x}_i^1 - \mathbf{x}_j^2\|^2}{\sigma_x^2} \right), & \text{if } \mathbf{x}_i^1 = \text{knn}(\mathbf{x}_j^2) \text{ and } \mathbf{x}_j^2 = \text{knn}(\mathbf{x}_i^1) \\ 0, & \text{otherwise} \end{cases}, \quad (17)$$

Note that, in Equation (17), we enforce the knn constraints using "and" instead of "or" to alleviate the possible false matching. Optimization of Equation (15) is built on the technique proposed in Sec. 3-A. We use Equation (9) to update $\mathbf{R}^1, \mathbf{R}^2$, and use Equation (11) to update $\mathbf{p}$. The extended formulation in Equation (15) shows that our proposed model generalizes both saliency and co-saliency detection.

C. Implementation Details

Our method is more sensitive to parameters than RPCA-based methods because of the non-convex formulation. Similar to [10], we propose a content-dependent approach to determine α, β, γ. We set $\alpha = \sigma_{\lfloor k/2 \rfloor}(\mathbf{X})$ for different images, and adaptively adjust β, γ at each iteration, where β is set as κ times of variance of $[\|\mathbf{x}_1 - \mathbf{r}_1\|^2, \ldots, \|\mathbf{x}_N - \mathbf{r}_N\|^2]^{\mathrm{T}}$ and $\gamma = \eta\beta$. In Sec. IV, we will discuss how to fine-tune the two additional parameters κ, η.

IV. Experimental Results

We compare our method with four recent state-of-the-art methods: HC/RC [16], HFT [21], ULRM [7], and MTSP [6]. Note that, all of these methods are published after 2011, and their reports showed that they outperform most of classic algorithms [13-15, 18-20, 22-24]. HC/RC utilizes global contrast measure on L*a*b color space and significantly outperforms local contrast-based methods. HFT clarifies the inappropriateness of spectral residual [18] and accordingly improves frequency-based methods. Both ULRM and MTSP rely on rank-sparsity decomposition. ULRM uses Equation (1) incorporated with segmentation and top-down priors, while MTSP uses Equation (2) incorporated with multi-task learning. We consider these four methods as the top of contrast-based, frequency-based, and uniqueness-based approaches.

We compare with HC/RC, HFT, and ULRM using their released code, and compare with MTSP using our own implementation because the code is not available. However, we find that the performance of MTSP is not as good as reported in [6]. For a fair comparison, we further use the average filter to smooth their saliency map for better performance. Finally, in the experiment of co-saliency detection, we compare with [28] using the author's released source code.

A. Datasets and settings

We conduct saliency detection experiments on the subset of MSRA [15] datasets. The MSRA subset is a salient-object dataset and contains 1000 images with binary ground truth. In addition, we conduct co-saliency detection experiments on the co-object dataset [29] containing 204 images with binary ground truth. To have a fair comparison with MTSP, we follow [6] to first resize each image into size 256x256 and then divide it into non-overlapping patches with size 8x8. We also adopt the same visual features, including 6 RGB color features, 13 steerable filter features, and 3 center-surround features, as suggested in [6]. Each feature vector $\mathbf{x}_i \in \mathbb{R}^{22}$ is then obtained from averaging all the pixels in the i^{th} patch. Fig. 1 demonstrates that the parameters κ and η controls the sparseness and the diversity of saliency scores. According to Fig. 1, we use $\kappa = 0.5$ and $\eta = 10$ for both two datasets. Finally, six commonly used numerical criteria are adopted for performance evaluation:

- **AUC**: Area under ROC curve, which is plotted by changing thresholds from 0 to 255 to obtain pairs of true positive rate (TPR) and false positive rate (FPR).
- **CCs**: Correlation coefficients between ground truth and saliency map (used in [6]).
- **MAE**: Mean absolute error between ground truth and saliency map (used in [17]). Note that we divide both the ground truth and saliency map by 255 to limit the range of MAE between 0 and 1.
- **Precision**, **Recall**, **F-measure**: Following [15], we use an adaptive threshold, which is defined as twice the mean saliency, for each image. Using the threshold, we determine the precision/recall, and then compute F-measure by

$$\text{F-measure} = \frac{(1+0.3) \times \text{Precision} \times \text{Recall}}{0.3 \times \text{Precision} + \text{Recall}}.$$

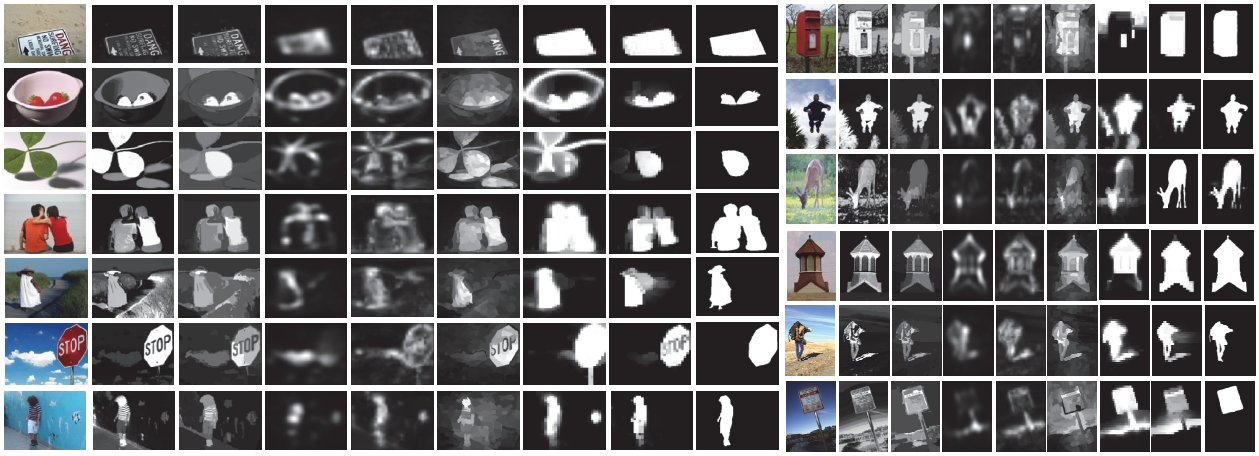
B. Results and discussion

1) Saliency detection

Table 1 and Fig. 2 show the quantitative and qualitative results in MSRA dataset. Our method IRS is superior to MTSP,

TABLE 1: AVERAGE OF SIX PERFORMANCE EVALUATIONS IN MSRA DATASET.

	HC	RC	HFT	MTSP	ULRM	IRS	$IRS_{Lab\text{-}cbias}$
AUC	0.9257	0.9467	0.9143	0.8702	0.9525	0.9184	0.9706
CCs	0.7073	0.7846	0.6278	0.5377	0.7695	0.6671	0.8499
MAE	0.1761	0.2272	0.1858	0.2463	0.2130	0.1354	0.0692
Precision	0.7418	0.8301	0.6841	0.6013	0.8537	0.6236	0.8246
Recall	0.7173	0.6034	0.6226	0.4741	0.6613	0.8180	0.8859
F-measure	0.7071	0.6824	0.6429	0.5257	0.7435	0.6451	0.8260



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

Figure 2: Some examples in MSRA dataset: (a) original images; (i) ground truth; and the saliency detection results obtained (b) HC; (c) RC; (d) HFT; (e) MTSP; (f) ULRM; (g) IRS; (h) $IRS_{Lab\text{-}cbias}$.

competitive with HC and HFT, but inferior to RC and ULRM. Note that, the setting of IRS is the same as MTSP (e.g., patch size, features). Since MTSP employs the rank-sparsity model based on Equation (2), the better performance indicates the superiority of our model over RPCA-based methods in saliency detection. On the other hand, RC and ULRM outperform our method because both of them utilize segmentation techniques. In addition, ULRM further includes two priors (color and location) while RC uses better discriminative color features (L*a*b rather than RGB). Because most salient objects in MSRA locate near image center and have discriminative colors, we conduct another experiment (denoted by $IRS_{Lab\text{-}cbias}$) by using only 6 color features obtained from L*a*b color space and including the proposed support of center-bias effect. From Table 1 and Fig. 2, we achieve the best performance over all state of the arts and obtain very consistent results with ground truth even without using any segmentation technique.

It is worth noting that our method tends to obtain smaller MAE, which is related to true negative rate (TNR), than all the other methods. As pointed in [17], although most approaches could produce acceptable saliency maps, they usually fail to discriminate non-salient pixels from salient pixels. In contrast, our method produces low TNR. This advantage can also be observed from Fig. 2, where contrast-based methods tend to obtain segment-like results while our method detects fewer non-salient pixels.

2) Co-saliency detection

We evaluate our co-saliency detection method proposed in Sec. 3-B (denoted by CoIRS) and also compare with IRS to show the benefits from co-saliency information. In addition, we validate their multi-task model extensions (denoted by IRS-MT and CoIRS-MT), and compare our results with a histogram-based co-saliency detection method [28].

Table 2 and Fig. 3 give the quantitative and qualitative results. In Table 2, CoIRS outperforms IRS, and their multi-task model extensions further improve the performance. These results are expected because now we include cross-image information to alleviate misestimates on non-salient regions. As shown in Fig. 3, IRS may fail to detect salient objects in cluttered background, but CoIRS successfully captures similar visual feature existing in both two images. On the other hand, [28] relies on histogram updates to capture similar visual cues but considers no spatial coherency. Therefore, although the co-saliency information is used, [28] still detects only small amount of co-salient pixels and obtains poor performance.

V. CONCLUSION

This paper proposes an implicit rank-sparsity model to improve existing rank-sparisty decomposition. Because our model considers the uncertainty of real-world observations, while applying to visual saliency detection, our method characterizes the implicit weighting factor between the underlying low-rank and sparse components as the saliency score. Because of its generality, we also extend our method to co-saliency detection without significant model changes. Experiment results demonstrate that the proposed method outperforms most state of the arts in two salient-object datasets;

TABLE 2: AVERAGE OF SIX PERFORMANCE EVALUATIONS IN CO-OBJECT DATASET.

	[28]	IRS	IRS-MT	CoIRS	CoIRS-MT
AUC	0.6188	0.8863	0.8911	0.9080	0.9137
CCs	0.1875	0.6068	0.6136	0.6403	0.6535
MAE	0.2489	0.2023	0.2039	0.1991	0.1993
Precision	0.5093	0.6280	0.6285	0.6606	0.6664
Recall	0.2295	0.6710	0.6763	0.6366	0.6452
F-measure	0.3621	0.6078	0.6075	0.6064	0.6097

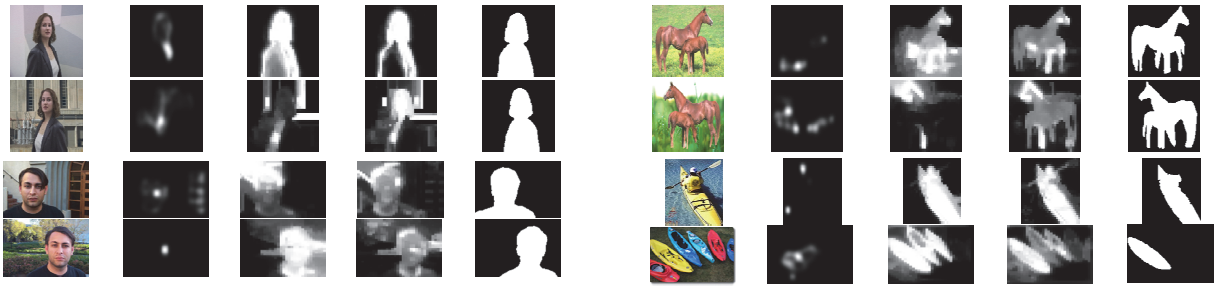


Figure 3: Some examples in co-object dataset: (a) original images; (e) ground truth; and the saliency detection results obtained by (b) [28]; (c) IRS; (d) CoIRS. The results of IRS-MT and CoIRS-MT are not shown here because they look very similar to (c)-(d).

and when using co-saliency information, we can further improve the detection performance.

REFERENCES

[1] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, "Rank-sparsity incoherence for matrix decomposition," *SIAM J. Optim.*, vol. 21, no. 2, pp. 572-596, 2011.

[2] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, "RASL: robust alignment by sparse and low-rank decomposition for linearly correlated images," in *Proc. CVPR*, 2010.

[3] Z. Zhang, A. Ganesh, X. Liang, and Y. Ma, "TILT: transform invariant low-rank texture," *Int. J. Computer Vision*, vol. 99, no. 1, pp.1-24, 2012.

[4] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, and Y. Ma, "Robust photometric stereo via low-rank matrix completion and recovery," in *Proc. ACCV,* 2010.

[5] J. Yan, M. Zhu, H. Liu, and Y. Liu, "Visual saliency detection via sparsity pursuit," *SPL*, vol. 17, no. 8, pp. 739-742, 2010.

[6] C. Lang, G. Liu, J. Yu, and S. Yan, "Saliency detection by multitask sparsity pursuit," *TIP*, vol. 21, no. 3, pp. 1327-1338, 2012.

[7] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Proc. CVPR*, 2012.

[8] R. Chartrand, "Nonconvex splitting for regularized low-rank+sparse decomposition," *TSP*, vol. 60, no. 11, pp. 5810-5819, 2012.

[9] X. Cui, Z. Huang, S. Zhang, and D. N. Metaxas, "Background subtraction using low rank and group sparsity constraints," in *Proc. ECCV*, 2012.

[10] X. Zhou, C. Yang, and W. Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation," *TPAMI*, vol. 35, no. 3, pp. 597-610, 2013.

[11] A. Borji, D. N. Sihite, and L. Itti, "Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study," *TIP*, vol. 22, no. 1, pp. 55-69, 2013.

[12] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *TPAMI*, vol. 35, no. 1, pp. 185-207, 2013.

[13] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *TPAMI*, vol. 20, no. 11, pp. 1254-1259, 1998.

[14] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *Proc. CVPR*, 2010.

[15] R. Achanta, S. S. Hemami, F. J. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. CVPR*, 2009.

[16] M. Cheng, G. Zhang, N. J. Mitra, X. Huang, and S. Hu, "Global contrast based salient region detection," in *Proc. CVPR*, 2011.

[17] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: contrast based filtering for salient region detection," in *Proc. CVPR*, 2012.

[18] X. Hou and L. Zhang, "Saliency detection: a spectral residual approach," in *Proc. CVPR*, 2007.

[19] C. Guo, Q. Ma, and L. Zhang, "Spatio-temporal saliency detection using phase spectrum of quaternion Fourier transform," in Proc. *CVPR*, 2008.

[20] X. Hou, J. Harel, and C. Koch, "Image signature: highlighting sparse salient regions," *TPAMI*, vol. 34, no. 1, pp. 194-201, 2012.

[21] J. Li, M. D. Levine, X. An, X. Xu, and H. He, "Visual saliency based on scale-space analysis in the frequency domain," *TPAMI*, vol. 35, no. 4, pp. 996-1010, 2013.

[22] N. Bruce and J. Tsotsos, "Saliency based on information maximization," in *Proc. NIPS*, 2006.

[23] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. NIPS*, 2007.

[24] X. Hou and L. Zhang, "Dynamic visual attention: searching for coding length increment," in *Proc. NIPS*, 2008.

[25] A. Borji and L. Itti, "Exploiting local and global patch rarities for saliency detection," in *Proc. CVPR*, 2012.

[26] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *Proc. ICCV*, 2009.

[27] J. Li, D. Xu, and W. Gao, "Removing label ambiguity in learning-based visual saliency," *TIP*, vol. 21, no. 4, pp. 1513-1525, 2012.

[28] H. Chen, "Preattentive co-saliency detection," in *Proc. ICIP*, 2010.

[29] H. Li, and K. N. Ngan, "A co-saliency model of image pairs," *TIP*, vol. 20, no. 12, pp. 3365-3375, 2011.

[30] J. Wright, A. Ganesh, S. Rao, and Y. Ma, "Robust principal component analysis: exact recovery of corrupted low-rank matrices via convex optimization," in Proc. *NIPS*, 2009.

[31] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recover of subspace structures by low-rank representation," *TPAMI*, vol. 35, no. 1, pp. 171-184, 2013.

[32] J. Cai, E. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956-1982, 2010.