From Co-saliency Detection to Object Co-segmentation: a Unified Multi-stage Low-rank Matrix Recovery Approach

Hao Chen, Panbing Wang, Ming Liu*

***Abstract*—Object co-segmentation aims to identify and segment the common objects among a set of similar images. Although various explorations have been done for the topic, two major problems still remain: (1) How to mitigate the influence of background disturbance of each image when we detect the common objects. (2) How to leverage common information of the image class optimally. To overcome the two problems, we resort to co-saliency detection and propose a novel framework, which utilizes multi-stage low-rank matrix recovery to eliminate the backgrounds and identify the common foregrounds. To address the first problem, we firstly use a conventional saliency detection model to get saliency map of each image as initialization rather than directly dealing with all the images together; to address the second problem, we adopt low-rank matrix recovery to constrain the common foregrounds as the low-rank part, while the background interferences correspond to the sparse noises. Besides, an effective refinement method is proposed to recover the spatial relationships among the segments. The extensive experiments show the proposed model can leverage the homogeneous information among the image class effectively and provide promising co-segmentation performance.**

I. Introduction

Saliency detection, which mimics the human visual attention mechanism for detecting what attracts humans the most [1], has been an important problem in computer vision.

Recently, with the proliferation of various photo-sharing websites, we note the shift that the image data are presented as a set which shares common information rather than individual samples. Aiming at identifying the common objects among multiple images, the co-saliency detection task has been a popular and challenging problem. Unlike traditional saliency detection models, which deals with individual image based on local contrast [1, 2], global contrast [3-5] or prior knowledge [6, 7] mainly, co-saliency detection additionally leverage the common information among multiple images. This technique will benefit various real-word applications such as robotic vision [8-11], image retrieval [12, 13] and video discovering [14]. Besides, co-segmentation, a closely related task to co-saliency, was introduced by Rother et al. [15]. It appends another related image to segment the common parts of an image pair. Using Markov Random Field (MRF), various models have been proposed by adding diverse constraints to optimize the cost functions [16-20]. Recently, the focus of co-segmentation has transferred to deal with multiple similar images rather than merely an image pair. It makes co-segmentation more applicable in computer vision. However, compared with co-saliency detection, co-segmentation has two particular features: firstly, without supervision information, co-segmentation cannot remove the common but no-salient regions effectively; secondly, some co-segmentation methods resort to human input [21] and interactive [22] to decide which cluster is the object. These methods are obviously not applicable to large-scale images due to high cost of manual labeling. Additionally, co-saliency detection can be readily used to achieve co-segmentation results of multiple images by appending an unsupervised befitting threshold. These differences enable the co-saliency detection can effectively remove the common background interference to generate better co-segmentation results. Meanwhile, it can mitigate the drawbacks of existing supervised co-segmentation methods. Therefore, we obtain the co-segmentation results via generating co-saliency maps firstly in this work.

Figure 1. To explain the co-saliency detection task. As shown in the first row, the footballer with yellow clothes in the first column, the footballer with green clothes and the referees in the second column, the footballer with blue clothes in the third column and the footballer with white clothes in the last column, they all should be salient in single saliency detection task, however, they turn into background interferences in the co-salient object detection task since only the footballers with red clothes are common salient objects.

As shown in Fig. 1, to obtain precise co-salient object segmentation, two keys problems should be solved: (1) How to mitigate the various background interferences of each image. (2) How to leverage common information of the image class optimally. Cluster-based methods are frequently adopted onto the detection of common foreground objects. However, it is hard to decide the number of clusters, especially when the background interferences are various. Besides, they always need to resort to prior knowledge or supervision information to decide the label association.

To address the two problems, we propose a unified multi-stage low-rank matrix recovery framework and obtain the co-segmentation results on the basis of co-saliency maps. As shown in Fig. 2. The first-stage low-rank matrix recovery is used to generate saliency map of each image; the second

Hao Chen and Panbing Wang are with the School of Automation, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: haochen593@gmail.com, wangpanbing@mail.nwpu.edu.cn).

Ming Liu is with the College of Science and Engineering, City University of Hong Kong (e-mail: mingliu@cityu.edu.hk).

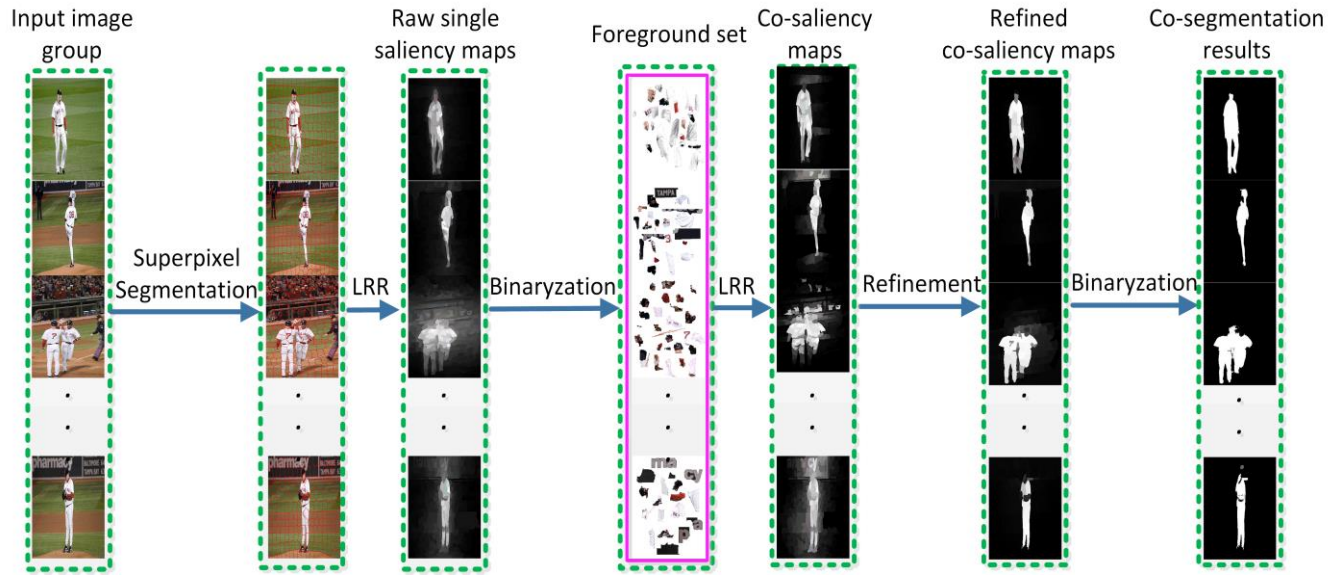

Figure 2. The framework of the proposed model.

one is adopted to constrain the common foreground with the low-rank structure. These operations simultaneously remove the background interferences from each image. Additionally, an effective refinement method is introduced to recover the spatial relationships among the super-pixels and to obtain a better uniformity and clearer boundary.

We would like to address the following contributions in this paper:

- Firstly, we leverage the advantages of co-saliency detection to achieve automatic co-segmentation without any supervision information, while it can achieve the state-of-the-art performance.
- Secondly, compared with widely-used cluster-based methods, the proposed method does not require inputting the number of clusters and judging which cluster corresponds to the common foreground regions, since the common foreground regions will be constrained as a lower-rank component through decomposition. The complex background interferences can be detected as the sparse part automatically.
- Thirdly, in the second stage low-rank matrix recovery, our method is insensitive to complex background. On the contrary, the more complex the backgrounds from multiple images are, the structure of the backgrounds in feature space will be sparser. Thus the low-rank part will be recovered more easily, which will benefit the separation of the common foregrounds and background interferences.

II. Related Work

A. Co-saliency detection

Co-saliency detection was introduced by Jacobs et al. in [23]. Aiming at getting co-salient regions between an image pair, Chen et al. [24] resorted to a pre-attentive scheme without performing the correspondence matching. Li et al. [25] linearly fused the saliency map on single image and co-saliency map via a complex co-multilayer graph. To extend the co-saliency detection from image pair to multiple images, Li et al. [26] proposed intra- and inter-saliency and combine them to obtain the co-saliency map. To leverage the global correspondence among multiple images, Fu et al. [27] utilized a cluster-based algorithm and multiply the saliency values obtained by different cues to generate the co-saliency map. Liu et al. [28] proposed a hierarchical segmentation based co-saliency model. On the basis of hierarchical segmentation, the global similarity of each region is derived based on regional similarity measures, and then intra-saliency map and object prior map are integrated to generate the co-saliency map for each image. Cao et al. [29] regarded the co-saliency detection task from another perspective, they combined the existing saliency detection models to generate a final co-saliency map, rather than resorting to the homogeneous information among the multiple related images. It is also worth mentioning that they also applied low-rank decomposition in their model. However, the low-rank decomposition was adopted to exploit the relationship of the saliency maps generated by different models to obtain the self-adaptive weights. Whereas in our model, we leverage low-rank matrix recovery to explore the low-rank structure of the common foregrounds among multiple images.

B. Co-segmentation

Rother et al. [15] introduced the definition of co-segmentation and proposed an energy model based on Markov Random Field (MRF), they leveraged histogram matching and a optimization method named trust region graph cuts (TRGC) to co-segment the common parts from an image pair. Following with the general MRF framework, several ideas made varied improvements and expansions from different aspects. L. Mukherjee et al. [18] replaced L1-norm distance employed in [15] with L2-norm distance to measure the foreground discrepancy and adopted quadratic pseudo Boolean optimization (QPBO) rather than TRGC to approach accurate results. Then D. S. Hochbaum et al. [19]

used a maximum flow procedure based on constructed graph model to solve this problem. S. Vicente et al. [20] made improvements on optimization by using Dual Decomposition techniques. K.-Y. Chang et al. [17] introduced the inconsistency between foreground and background of each image instead of simply taking the foreground coherence into consideration. J. C. Rubio et al. [16] structured the energy function based on MRF through region matching via establishing correspondences between the common objects to explore inter-image information.

Unlike the mentioned methods which are based on MRF and mostly dealing with an image pair, A. Joulin et al. [30] implemented co-segmentation via a classifier trained by integrating discriminating clustering with spectral clustering and successfully expanded it to multi-class task through energy minimization [31]. However, it required human labor to choose which cluster was more possible to be foreground. Aiming at identifying which cluster corresponded to the objects, supervision information was added with various strategies. As mentioned in [18] and [19], some scribbles should be suggested firstly as the guidance for choosing the cluster which contains the objects. As a substitute, D. Batra et al [22] proposed to leverage manual strokes on the region which was the hardest to judge its pixel labels, and J. Cui et al. [32] introduced a premise that one of the images was hand-segmented. Without adding supervision information or interactive labor, Yong Li et al. [33] composited low-rank matrix recovery with discriminating learning to realize object co-segmentation. However, this method directly input all the similar images without identifying the salient regions of each image firstly, which will significantly decrease the effectiveness when the background interferences are common, since the structure of the common background interferences in feature space will also lie in a low-rank subspace.

III. The Proposed Method

The proposed model as shown in Fig. 2 consists three parts: foreground queries generation, co-salient object detection via low-rank matrix recovery and spatial correlation recovery.

A. Foreground queries generation

We follow [34] to get the saliency map of each image and generate the foreground queries with an adaptive threshold, which can be seen as the initiation of our object co-segmentation model. For the *i*-th image, we can get N_i superpixels via mean-shift algorithm [35]. (We set N_i=200 empirically). Therefore, the feature representation of the *i*-th image can be noted as $\mathbf{F}_i = [f_1 \ f_2 \ \cdots \ f_j \ \cdots \ f_N]$, where f_j is the feature representation of the *j*-th superpixel. The feature representation of each superpixel f_j is the mean of the features in this segment, and $f_j \in \mathbf{R}^{\mathbf{V}\times 1}$, where $\mathbf{V}$ =53 denotes the dimension of the adopted features which contains color appearance and textural features, amount to 53 dimensions, we suggest readers to refer to [34] for details. We indicate the low-rank matrix and sparse noises of the *i*-th image with $\mathbf{L}_i$ and $\mathbf{S}_i$ respectively. The image can be represented as $\mathbf{F}_i = \mathbf{L}_i + \mathbf{S}_i$, where the low-rank part corresponds to the background regions while the sparse noise indicates the salient regions. Following this idea, we can model the problem by

$$(\mathbf{L}_i^*, \mathbf{S}_i^*) = \arg\min \left(\|\mathbf{L}_i\|_* + \lambda_1 \|\mathbf{S}_i\|_1 \right) \quad s.t. \mathbf{F}_i = \mathbf{L}_i + \mathbf{S}_i \qquad (1)$$

where $\|\mathbf{L}_i\|_*$ expresses the nuclear norm of $\mathbf{L}_i$ which restricts the rank of $\mathbf{L}_i$ and $\|\mathbf{S}_i\|_1$ represents the l_1-norm of $\mathbf{S}_i$, and the l_1-norm of each column S_{ij} in $\mathbf{S}_i$ is used to measure the saliency values of the corresponding superpixels. A larger $\|S_{ij}\|_1$ means a greater likelihood of the *j*-th superpixel to be salient and results in a greater assigned saliency value. Then we accordingly generate the saliency map and normalize it to be a gray-scale image. Here λ_1 is used to tradeoff between the $\mathbf{L}_i$ and $\mathbf{S}_i$ .We follow [35] to set all the parameters without further tuning.

Besides, high-level priors which contain location, semantic and warm color cues are added to enhance the saliency detection results, and a feature transformation matrix $\mathbf{T}$ is learned to generate a good feature space, in which most image background should lie in a subspace with low dimensional so that they can be represented as a low rank matrix. Then the input matrix will be decomposed to $\mathbf{TFP}_i = \mathbf{T} * \mathbf{F}_i * \mathbf{P}$, where $\mathbf{P}$ denotes the prior matrix. Therefore Eq. (1) is advanced to

$$(\mathbf{L}_i^*, \mathbf{S}_i^*) = \arg\min \left(\|\mathbf{L}_i\|_* + \lambda_1 \|\mathbf{S}_i\|_1 \right) \ s.t. \mathbf{TFP}_i = \mathbf{L}_i + \mathbf{S}_i \qquad (2)$$

To make exact recovery of corrupted low-rank matrices, Augmented Lagrange Multiplier (ALM) method [36] is adopted to optimize the formulation. After this process, we set the threshold $T_i = 1.5 * mean(\mathbf{S}_i)$ to select the foreground queries in each image.

By this way, we can choose the foreground superpixels of all the images in the same class, and then we represent the features of the foreground queries of the *i*-th image by:

$$\mathbf{I}_i = [\Omega_1^i \ \Omega_2^i \ldots \Omega_k^i \ \ldots \ \Omega_m^i] \qquad (3)$$

where Ω_k^i represents the features of the *k*-th foreground superpixel in the *i*-th image and *m* denotes the number of the foreground queries of the *i*-th image. Therefore, the feature matrix of the collected foreground queries of the *N* images of an image class can be represented as $\mathbf{I} = [\mathbf{I}_1 \ \mathbf{I}_2 \ \ldots \ \mathbf{I}_i \ \ldots \ \mathbf{I}_N]$.

B. Co-salient object detection

After we have obtained the foreground queries of each image, we use the second-stage low-rank matrix recovery to obtain co-salient regions by inputting all the foreground queries together. Since the similar salient regions share common patterns, the common salient regions will lie in low-rank structure, while the sparse part can be explained as the background interferences from different images. This strategy can effectively exclude the background interferences among the foreground candidates.

As a result, the feature matrix of coarse foreground set can be represented as $\mathbf{I} = \mathbf{L}_I + \mathbf{S}_I$, where the low-rank matrix

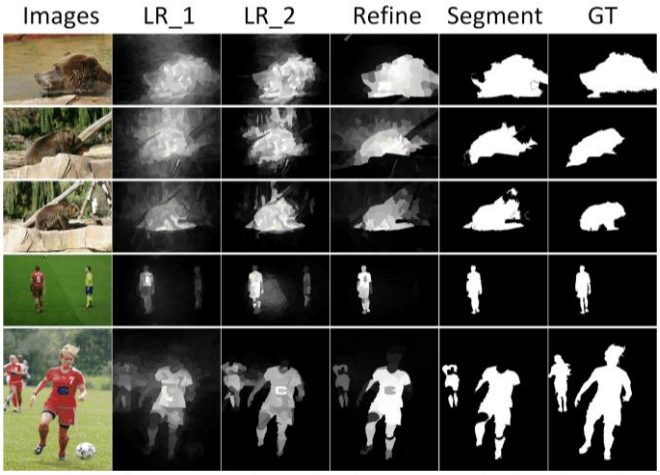

Figure 3. Qualitative comparisons of each component, the six columns from left to right successively denote the input images, the saliency maps after the first-stage low-rank matrix recovery, the coarse co-saliency maps after the second-stage low-rank matrix recovery, the co-saliency maps after graph based refinement, the eventual co-segmentation maps, and the groundtruth.

$\mathbf{L}_I$ corresponds to the common foregrounds while the sparse matrix $\mathbf{S}_I$ indicates the background interferences. Following this idea, we can model the problem as

$$(\mathbf{L}_I^*, \mathbf{S}_I^*) = \arg\min\left(\|\mathbf{L}_I\|_* + \lambda_I \|\mathbf{S}_I\|_1\right) \quad s.t. \quad \mathbf{I} = \mathbf{L}_I + \mathbf{S}_I \tag{4}$$

where $\mathbf{L}_I$ and $\mathbf{S}_I$ can be perfectly recovered by ALM [36].

$\mathbf{S}_I$ contains the sparse parts of the N images, that is

$$\mathbf{S}_I = [S_1^{co} \;\; S_2^{co} \;\; \ldots \;\; S_i^{co} \;\; \ldots \;\; S_N^{co}] \tag{5}$$

After we have obtained $\mathbf{L}_I$ and $\mathbf{S}_I$, the l_1-norm of each column s_{ik}^{co} in s_i^{co} is used to measure the saliency of corresponding segments. If $\left\|s_{ik}^{co}\right\|_1$ is smaller, we assign a higher saliency value to the k-th segment. A coarse co-saliency map ψ_{co}^i is then accordingly generated and normalized to be a gray-scale image.

C. Spatial correlation recovery

When we stack all the foreground queries to generate the coarse co-saliency maps via the second-stage low-rank matrix recovery, spatial correlation among the superpixels in each image is totally broken and ignored. However, the spatial-wise coherence is another important cue in addition to appearance for a salient object, so we adopt a graph-based model to recover the spatial relationships among the superpixels. For each image, we follow [6] to establish a graph model G=(V, E) with the superpixels as nodes,. The edges E are weighted by an affinity matrix $\mathbf{W} = [w_{ij}]_{N_i \times N_i}$, where the weight between two connected nodes w_{ij} is measured by their feature distance (i.e. the Euclidean distance of the mean values in the CIT LAB color space) as defined in [6]. The degree matrix $\mathbf{D}$ for the given G is $\mathbf{D} = \mathrm{diag}\left\{d_{11}, \ldots, d_{N_i N_i}\right\}$, where $d_{ii} = \sum_j w_{ij}$.

Figure 4. Co-segmentation maps to show the effectiveness of our model.

The completed affinity matrix becomes $(\mathbf{D} - \partial * \mathbf{W})^{-1}$ and the refined saliency map can be formed by

$$S_{re}^i = (\mathbf{D} - \partial * \mathbf{W})^{-1} * \psi_{co}^i \tag{6}$$

Then we segment the refined co-saliency maps by an adaptive threshold (e.g. the twice of the mean value of the co-saliency map) to generate the object co-segmentation results.

IV. Experiments

A. Evaluation metrics

We evaluate the performance of our method on various image classes via the parameter average intersection-over-union (IOU) score:

$$\mathbf{IOU} = \frac{1}{\tau} \sum_{i \in \tau} \frac{R_i \cap GT_i}{R_i \cup GT_i} \tag{7}$$

where R_i is the co-segmentation result and GT_i indicates the corresponding groundtruth mask. τ denotes the number of images of an image class sharing common objects.

B. Comparisons with different components

We evaluate our model on the widely-used iCoseg dataset [22], which consists 38 classes amount to 643 images.

We generate the saliency maps of the three key stages respectively to verify the contributions of each component. As shown in Fig. 3, compared with the saliency maps obtained by the first-stage low-rank matrix recovery (LR_1), the co-saliency maps generated after the second-stage low-rank matrix recovery (LR_2) can effectively highlight the co-salient regions and remove the background interferences. And the spatial correlation recovery stage (denotes as 'Refine') can leverage the spatial relationships among the superpixels to refine the co-saliency maps effectively. The refined co-saliency maps supply clearer and more exact edges between foreground and background. Additionally, the saliency values are more uniform after the refinement stage.

Moreover, as shown in Fig. 4, our model can exactly eliminate the salient but not common regions in each image (e.g. the white football player) while identifying the common salient object (e.g. the red football player) precisely.

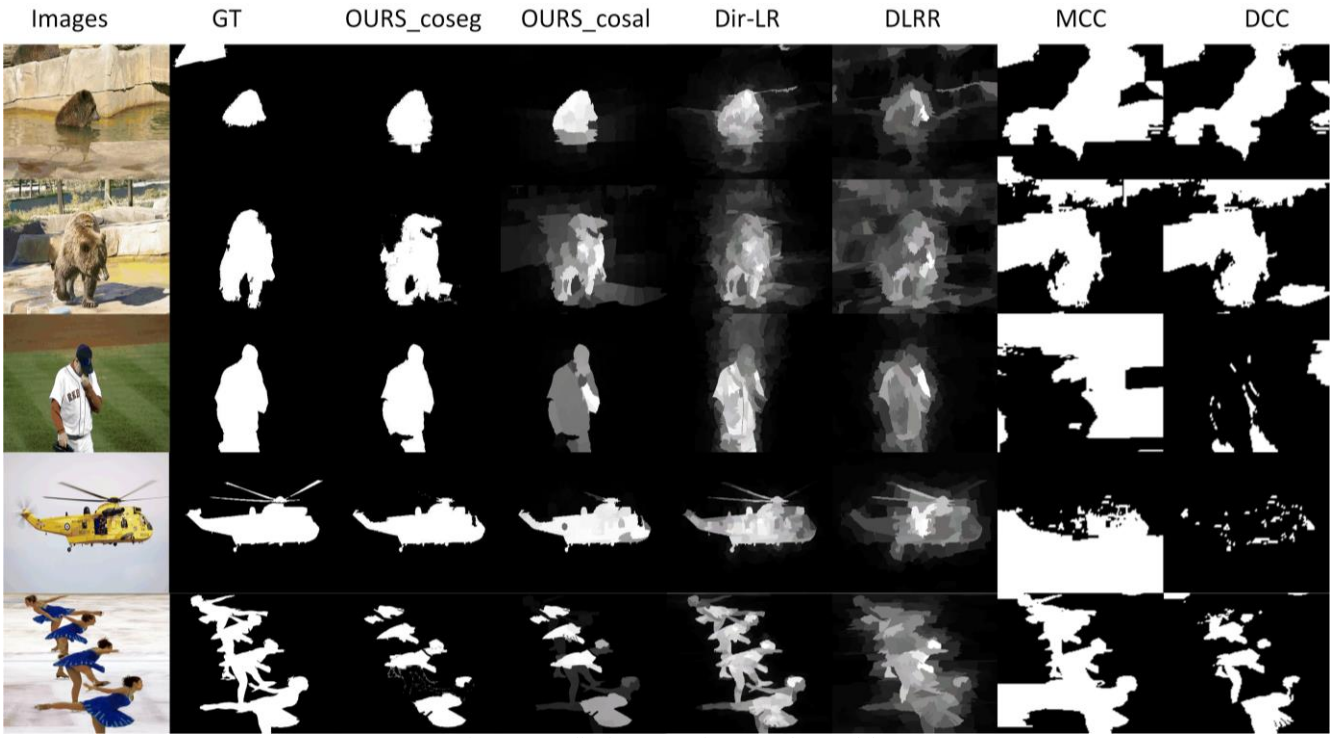

Figure 5. Qualitative comparisons to other models.

TABLE I. QUANTITATIVE COMPARISONS WITH OTHER MODELS ON THE ICOSEG DATASET

CLASS	OURS	Dir-LR	DLRR	MCC	DCC
002 AlaskanBear	**0.63102**	0.4699	0.4108	41.6	46.1
006 Red Sox Players	**0.64626**	0.5004	0.6288	13.6	31.4
014 Liverpool	**0.52849**	0.365	0.3931	38.7	14.9
015 Ferrari	**0.69972**	0.4888	0.5457	38.7	26.4
017 Taj Mahal	0.453941	0.4002	0.4665	37.1	38.4
025 helicopter	**0.77072**	0.7603	0.6235	33.3	61
032 Kite-Brighton	**0.87737**	0.6747	0.4591	22.1	57.8
037 Skating-Skating	0.412381	0.5646	0.5419	72.7	38.1
040 Monks	0.661772	0.4765	0.4037	73.8	68.4
brown_bear	0.562828	0.3949	0.3956	57.5	49.4
AVG(10 classes)	**0.62445**	**0.50953**	**0.48687**	**42.91**	**43.19**
AVG(38 classes)	**0.59772**	**0.4683**		**0.3762**	**0.4144**

The comparison of IOU values with other models, since the DLRR model only public the results of 10 classes, we choose the corresponding results of other models to compare. Besides, the average value of totally 38 classes are also shown.

C. Comparisons to other models

To evaluate the effectiveness of the proposed model and confirm the validity of each step of our method, we compare our model with DLRR [36], MCC [31], and DCC [30]. To confirm the effectiveness of the proposed co-salient object detection stage, we also compare our results with the single saliency pursuing stage, which is indicated by Dir-LR [34]. From Table I we can see that our model outperforms others obviously, we can also conclude that the eventual result contains less background regions compared with other models shown in Fig. 5. Additionally, comparing our model with Dir-LR, we can conclude that the proposed co-salient object detection strategy with low-rank matrix recovery can effectively extract the common salient regions among the image class.

D. Discussions

In the stage of co-salient regions detection stage of our model, we deem the background interferences from different images as sparse part with the assumption that the background interferences are diverse in feature space or they should account a relative small part in the collected coarse foreground set. The reasons are as follows: low-rank is a relative relationship, no explicit stipulation has been proposed to state that what size of the rank should be regard as low-rank. On the condition that the background interferences are consistent and occupy a relative competitive size, the background interferences can also share a low-rank structure. Therefore, the proposed co-salient object detection via low-rank matrix recovery is more effective when the background regions are more complex and diverse since it is harder for the background regions to share a low-rank structure on this occasion.

Since our model highly depends on the initiation stage to obtain the individual saliency maps, it is obvious that different conventional saliency detection methods can make disparate contributions to the final results. Thus the co-saliency detection results will be improved by enhancing the saliency detection on single image in the future. Besides, with the development of 3D technology, depth information such as point cloud has contribute a lot to vision tasks [37, 38]. Therefore, resorting to depth information to enhance the co-segmentation will be another focus of us in the future.

V. CONCLUSIONS

In this paper, we proposed a new model to obtain object co-segmentation via co-saliency detection based on a unified multi-stage low-rank matrix recovery framework. The proposed model can identify the common foregrounds, and eliminate the background interferences effectively. Besides, an effective refinement method is proposed to recover the spatial relationships among the segments. The experiments showed the effectiveness of the proposed model and verified the contributions of each step of our method.

VI. ACKNOWLEDGEMENT

This work is supported by the Research Grant Council of Hong Kong SAR Government, China, under project No. 16206014 and No. 16212815; National Natural Science Foundation of China No. 6140021318, awarded to Prof. Ming Liu.

REFERENCES

[1] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, *et al.*, "Learning to detect a salient object," *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, vol. 33, 2011, pp. 353-367.

[2] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, 1998, pp. 1254-1259.

[3] M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S. Hu, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, 2015, vol. 37, pp. 569-582.

[4] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Computer Vision and Pattern Recognition. IEEE Conference on*, 2012, pp. 733-740.

[5] P. Jiang, H. Ling, J. Yu, and J. Peng, "Salient region detection by ufo: Uniqueness, focusness and objectness," in *Computer Vision. IEEE International Conference on*, 2013, pp. 1976-1983.

[6] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Computer Vision and Pattern Recognition. IEEE Conference on*, 2013, pp. 3166-3173.

[7] B. Alexe, T. Deselaers, and V. Ferrari, "What is an object?," in *Computer Vision and Pattern Recognition. IEEE Conference on*, 2010, pp. 73-80.

[8] M. Liu, K. Qiu, F. Che, S. Li, B. Hussain, L. Wu, *et al.*, "Towards indoor localization using Visible Light Communication for consumer electronic devices," in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, 2014, pp. 143-148.

[9] M. Liu and R. Siegwart, "Topological mapping and scene recognition with lightweight color descriptors for an omnidirectional camera," *Robotics, IEEE Transactions on*, vol. 30, pp. 310-324, 2014.

[10] M. Liu, F. Colas, and R. Siegwart, "Regional topological segmentation based on mutual information graphs," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, 2011, pp. 3269-3274.

[11] M. Liu and R. Siegwart, "DP-FACT: Towards topological mapping and scene recognition with color for omnidirectional camera," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, 2012, pp. 3503-3508.

[12] A. Toshev, J. Shi, and K. Daniilidis, "Image matching via saliency region correspondences," in *Computer Vision and Pattern Recognition. IEEE Conference on*, 2007, pp. 1-8.

[13] J. Xue, L. Wang, N. Zheng, and G. Hua, "Automatic salient object extraction with contextual cue and its applications to recognition and alpha matting," *Pattern Recognition*, vol. 46, 2013, pp. 2874-2889.

[14] J. Guo, Z. Li, L.-F. Cheong, and S. Z. Zhou, "Video co-segmentation for meaningful action extraction," in *Computer Vision. IEEE International Conference on*, 2013, pp. 2232-2239.

[15] C. Rother, T. Minka, A. Blake, and V. Kolmogorov, "Cosegmentation of image pairs by histogram matching-incorporating a global constraint into mrfs," in *Computer Vision and Pattern Recognition. IEEE Computer Society Conference on*, 2006, pp. 993-1000.

[16] J. C. Rubio, J. Serrat, A. López, and N. Paragios, "Unsupervised co-segmentation through region matching," in *Computer Vision and Pattern Recognition. IEEE Conference on*, 2012, pp. 749-756.

[17] K.-Y. Chang, T.-L. Liu, and S.-H. Lai, "From co-saliency to co-segmentation: An efficient and fully unsupervised energy minimization model," in *Computer Vision and Pattern Recognition. IEEE Conference on*, 2011, pp. 2129-2136.

[18] L. Mukherjee, V. Singh, and C. R. Dyer, "Half-integrality based algorithms for cosegmentation of images," in *Computer Vision and Pattern Recognition. IEEE Conference on*, 2009, pp. 2028-2035.

[19] D. S. Hochbaum and V. Singh, "An efficient algorithm for co-segmentation," in *Computer Vision IEEE 12th International Conference on*, 2009, pp. 269-276.

[20] S. Vicente, V. Kolmogorov, and C. Rother, "Cosegmentation revisited: Models and optimization," in *Computer Vision–ECCV 2010*, ed: Springer, 2010, pp. 465-479.

[21] D. Kuettel and V. Ferrari, "Figure-ground segmentation by transferring window masks," in *Computer Vision and Pattern Recognition. IEEE Conference on*, 2012, pp. 558-565.

[22] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen, "icoseg: Interactive co-segmentation with intelligent scribble guidance," in *Computer Vision and Pattern Recognition. IEEE Conference on*, 2010, pp. 3169-3176.

[23] D. E. Jacobs, D. B. Goldman, and E. Shechtman, "Cosaliency: Where people look when comparing images," in *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, 2010, pp. 219-228.

[24] H.-T. Chen, "Preattentive co-saliency detection," in *Image Processin. 17th IEEE International Conference on*, 2010, pp. 1117-1120.

[25] H. Li and K. N. Ngan, "A co-saliency model of image pairs," *IEEE Transactions on Image Processing.*, vol. 20, 2011, pp. 3365-3375.

[26] H. Li, F. Meng, and K. N. Ngan, "Co-salient object detection from multiple images,", *IEEE Transactions on Multimedia.*, vol. 15, 2013, pp. 1896-1909,.

[27] H. Fu, X. Cao, and Z. Tu, "Cluster-based co-saliency detection," *IEEE Transactions on Image Processing.,2013*, vol. 22, pp. 3766-3778.

[28] Z. Liu, W. Zou, L. Li, L. Shen, and O. Le Meur, "Co-saliency detection based on hierarchical segmentation," *Signal Processing Letters, IEEE*, vol. 21, 2014, pp. 88-92.

[29] X. Cao, Z. Tao, B. Zhang, H. Fu, and W. Feng, "Self-Adaptively Weighted Co-Saliency Detection via Rank Constraint," *IEEE Transactions on Image Processing.*, vol. 23, 2014, pp. 4175-4186.

[30] A. Joulin, F. Bach, and J. Ponce, "Discriminative clustering for image co-segmentation," in *Computer Vision and Pattern Recognition. IEEE Conference on*, 2010, pp. 1943-1950.

[31] A. Joulin, F. Bach, and J. Ponce, "Multi-class cosegmentation," in *Computer Vision and Pattern Recognition. IEEE Conference on*, 2012, pp. 542-549.

[32] J. Cui, Q. Yang, F. Wen, Q. Wu, C. Zhang, L. Van Gool, *et al.*, "Transductive object cutout," in *Computer Vision and Pattern Recognition. IEEE Conference on*, 2008, pp. 1-8.

[33] H. Li, F. Meng, Q. Wu, and B. Luo, "Unsupervised Multi-class Region Co-segmentation via Ensemble Clustering and Energy Minimization," *IEEE Transactions on Circuits and Systems for Video Technology*., vol. 24, 2014, pp: 789-801.

[34] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Computer Vision and Pattern Recognition. IEEE Conference on*, 2012, pp. 853-860.

[35] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, vol. 24, 2002, pp. 603-619.

[36] Y. Li, J. Liu, Z. Li, Y. Liu, and H. Lu, "Object co-segmentation via discriminative low rank matrix recovery," in *Proceedings of the 21st ACM international conference on Multimedia*, 2013, pp. 749-752.

[37] M. Liu, "Efficient Segmentation and Plane Modeling of point-cloud for structured environment by Normal Clustering and Tensor Voting." *Robotics and Biomimetics (ROBIO), IEEE International Conference on, 2014*, pp. 1805 - 1810.

[38] M. Liu, C. Pradalier, Q. Chen, and R. Siegwart, "A bearing-only 2D/3D-homing method under a visual servoing framework," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, 2010, pp. 4062-4067.