ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 12, July 2017

Object Co-segmentation for Irrelevant Video Frames Involved in Multimedia

Ayesha Firdose[1], Dr. Aravinda T V[2]

P.G. Student, Department of CSE, SJMIT, Chitradurga, India[1]

Professor & PG Coordinator, Department of CSE, SJMIT, Chitradurga India [2]

ABSTRACT*:* Despite the fact that there have been a lot of past work on video segmentation technique, it is still a difficult task to extricate the video objects precisely without interactions, particularly for those videos which contain irrelevant frames (frames containing no normal targets). In this paper, a novel multi-video object co-segmentation technique is raised to co-segment normal or comparable objects of pertinent casings in various videos, which incorporates three stages: 1) object proposal generation and clustering within each video; 2) weighted graph construction and common objects selection; and 3) irrelevant frames detection and pixel-level segmentation refinement.

KEYWORDS: Irrelevant frames detection, loopy belief *propagation*, proposal, video co-segmentation.

I. INTRODUCTION

Video object segmentation has played an important role in senior computer vision tasks, such as activity recognition, content-based retrieval, and object tracking. However, due to the complicated spatial-temporal correlations, large data, and high scene complexity, video object segmentation still faces serious challenges

Recently, to make full use of the additional information across multiple videos, video co-segmentation [9][17] has become another research hotspot. Researchers propose to use the information across multiple videos to realize the segmentation of the common targets. The methods in [9][11] enforce the cooperative constraint with a global appearance model for the common targets. In [12] and [13], the trajectory cosaliency constraint and the coherent moving constraint for the foreground local parts are introduced into video co-segmentation, respectively. And more recently, the object-based approaches in [14][17] discover the common targets by mining the consistency information among the segmentation proposals. The cooperative constraint brings more information, while a neglected problem is that the interference information brought by the noisy frames, which contain no common targets, makes the video co-segmentation more complicated. Actually, these irrelevant frames involved videos are very common, such as a target is out of the view or occluded by the background as presented in Fig. 1.

In this letter, a more direct and unsupervised method is proposed, which is based on common cluster selection and adaptive recognition for the irrelative frames. The method has two key contributions. First, a unified framework for multi-video common targets' mining is proposed. Proposed system construct a weighted graph which regards each object cluster as a graph node and takes the feature similarities from every pair of object clusters as edge weights. Then selection of common object clusters turns into searching the edge set with maximum weight, which can be solved with A^*-search algorithm [19] or loopy belief propagation for approximation. Second, with the chosen object clusters, this paper introduces an adaptive decision condition to detect the irrelevant frames automatically. Experiments on two public datasets demonstrate the good performance of the proposed method.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 12, July 2017

Fig. 1. Examples of irrelevant frame-involved videos. The first two rows show the objects' balloon and bear go into or disappear from the scenes as time passes, and the third row shows the situation that the foreground is occluded, here the Ferrari car is completely blocked by the woods.

II. Our Approach

This paper presents an overview of the proposed approach in Fig. 2(a). And accordingly, it will introduce the proposed method in this section from the following three stages.

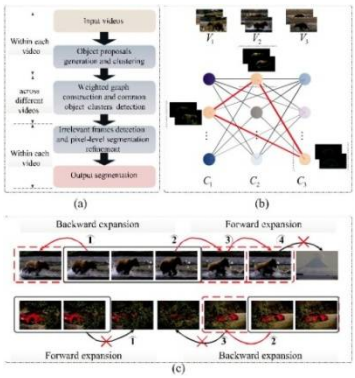

Fig. 2. Algorithm flowchart and the key steps. (a) Overall framework of the proposed video co-segmentation method. (b) Weighted graph construction with proposal clusters from multiple videos and common object clusters' detection. The proposal clusters (denoted with solid nodes) of the same video are arranged in the same column. The nodes of different videos are linked with edges in different thicknesses, whose weights correspond to the possibility of the con-nected clusters being both real common targets. For clarity, only part of the weighted edges is shown. Particularly, the nodes connected by red lines are the selected common object clusters. (c) Examples of chosen cluster expansion and irrelevant frames detection. The original chosen clusters are in black boxes and the newly expanded frames are in the red boxes. The numbers indicate the expansion order.

A. Intravideo Object Proposal Generating and Clustering

Let $V = \{V_i\}$, $i = 1, 2, \ldots, N$ be the original video set, where video $V_i = \{f_i^j\}$, $j = 1, 2, \ldots, M_i$, M_i is the number of the frames contained in video V_i. the proposed system first generate an Object proposal set $P_i = \{p^1_{i,1}, p^1_{i,2}, \ldots, p^1_{i,K1}, p^2_{i,1}, p^2_{i,2}, \ldots, {}_{K2}, \ldots, p^{M_i}_{i,1}, p^{M_i}_{i,2}, \ldots, p^{M_i}_{i,K\,Mi}\}$ with [20] for each video V_i, where K_j is the number of proposals of jth frame.To measure the possibilities of those proposals to be real objects, it utilizes a score function which combines appearance and motion information

$$S(p) = S_a(p) + S_m(p) \qquad (1)$$

where p is the object proposal to be measured, $S_a(p)$ is the appearance score that measures the boundary occlusion of proposal as well as the appearance difference from its nearby pixels. The appearance score $S_a(p)$ is computed as defined in [20], which is achieved with a pretrained ranker. Additionally, to measure the probability of a region belonging to a moving object, it also computes the motion score as defined in [5]

$$S_m(p) = 1 - \exp -\chi^2_{flow}(p, p^{-}) \qquad (2)$$

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 12, July 2017

It measures the χ^2 distance of optical flow histograms for proposal region p and its surrounding pixels $\bar{p}$ within the loosely fitting bounding box. According to $S(p)$, top 20 proposals of each frame are selected as the target candidates, which are most likely to be the true targets.

B. Inter-video Common Objects Selection

The proposal clusters achieved by the previous stage usually describe different targets, even some object-like backgrounds. Video co-segmentation aims at mining the common objects shared by all the videos, so here this paper propose an objects clusters' selection method to pick out the desired common targets automatically. As presented in Fig. 2(b), a weighted graph is constructed with each node corresponding to a proposal cluster. The nodes presenting the clusters of the same video are arranged in the same column. To establish global relationships for multiple videos, it connects every two nodes of different videos. The weights of the edges are formulated as

$$G(c_u,c_v) = G_c(c_u,c_v)+\lambda G_s(c_u,c_v) \quad (3)$$

where c_u and c_v are two clusters belonging to different videos, $G_c(c_u, c_v)$ is their color similarity that reflects the consistency, $G_s(c_u, c_v)$ is their saliency score that measures the unary target likelihood, λ is the weight coefficient which is fixed to 0.5 in the experiments. These two terms are defined as follows:

$$G_c(c_u,c_v) = \sum_{\alpha\text{-}1}^{\text{length(h)}} \min(h_u(\alpha),h_v(\alpha)) \quad (4)$$

where h_u and h_v are color histograms of c_u and c_v in Lab color space, here it only keep channels a and b for illumination invariance.

$$G_s(c_u,c_v) = \sum_{p\epsilon c_u} S(p)/N_{CU} + \sum_{q\epsilon c_v} S(q)/N_{cv} \quad (5)$$

The object clusters selection means discovering the common targets shared by the video set, which should be close in feature space. And then, the clusters selection task turns into finding the nodes with the largest weight sum, i.e., giving each cluster a label $x \in \{0, 1\}$ (1 for target, 0 for nontarget) by maximizing the following function:

$$E(x) = \sum_{m\le N_{Cu}, n\le N_{Cu}} G(x^m_u, x^n_v), \text{ s.t.} \sum_{J=1}^{N_{ci}} c_i x^j_i = 1 \quad (6)$$

where $G(x^m_u, x^n_v) = G(c^m_u, c^n_v)$ if $x^m_u = x^n_v = 1$ or $G(x^m_u, x^n_v) = 0$, otherwise. This problem can be solved efficiently with loopy belief prorogation.

C. Chosen Cluster Expansion and Segmentation Refinement

As the illumination or position changes, the chosen object cluster may omit several object proposals whose similarities with cluster center are relatively low. More generally, common targets may not always exist throughout the whole video. In this section, a chosen proposal cluster expansion mechanism is raised to adaptively retrieve the omissive target proposals as well as to delete those irrelevant frames. Within each video, the proposed system observes that if the proposals in omissive frames (with no proposals contained in the chosen cluster) really describe the desired targets, they should be relatively close with the proposals of chosen cluster in feature space. In the experiments. The detailed definition of the similarity measurements is as follows:

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 12, July 2017

$$Sim_c(p,c_u) = \sum_{p_i \epsilon c_u} G_r(p,p_i) / N_{cu} \quad (7)$$

where $G_r(p, p_i)$ measures the color similarity between p and p_i in RGB color space as (4) describes

$$Sim_c(c_u) = \sum_{p_i, p_j \epsilon c_u} G_r(p_i,p_j) / \binom{N_{cu}}{2} \quad (8)$$

In Fig. 2(c), the proposed system presents two examples to illustrate the chosen cluster expansion. Before expansion, it first labels all the frames whose proposals are included in c_u by 1. Then, it start detecting the irrelevant frames according to the sequence distance between the chosen frames and those missed frames, such as the series number in Fig. 2(c). Once the condition in 7 is satisfied by a proposal p of f_{miss}, it will add the proposal p into the chosen cluster and label f_{miss} as 1. Otherwise, f_{miss} will be determined as an irrelevant frame and labeled as 0. After achieving the expanded chosen clusters, to further refine the target segmentations, it generates foreground/background models with Gaussian mixture models (GMM) .

III. RESULTS

To verify the performance of the proposed approach, the proposed system evaluates the method on MOViCS dataset [11] as well as video object co-segmentation dataset [18] which contains irrelevant frames. The proposed system employs two different evaluation metrics on the datasets, respectively. For the MOViCS dataset, it utilizes intersection-over union metric, and for the latter, the proposed system uses the number of misjudged frames to verify the effectiveness and accuracy of the method.

A. MOViCS Datasets

MOViCS [11] contains 11 recordings in 4 groups, and the proposed system labels ground truth for every frame. In the method, it only consider the object which belongs to the same class and appears in all the videos. Particularly, for all the four lion–zebra videos, it picks three recordings which have lions in common.It retests video segmentation methods [7], [17] with the public code under the same setting and revaluates them with all the frames. Fig. 3 shows several visualization results and the quantitative comparison is given in Table I. Here, it also reports the upper bound of the performance of ([20]-Upper) and the intermediate results after cluster expansion (Ours-Expand). It can be noticed that, compared with single video segmentation method[7] and cosaliency method , the method performs better since it gives a global comprehension of the video set when it contains disturbance

Fig. 3. Segmentation examples of the proposed method and two comparison methods [7], [17]. In each sheet, the first column presents the original video frames; from the second to the fourth columns, the segmentations of [7] and [17] and our method are displayed, respectively.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 12, July 2017

TABLE 1
QUANTITATIVE SEGMENTATION RESULTS OF THE PROPOSED METHOD APPROACHES ON MOVICS DATASET

Video set	Chicken	Giraffe	Lion	Tiger	Average
CoSal[22]	0.53	0.36	0.32	0.41	0.40
LDAG[7]	0.57	0.35	0.57	0.65	0.55
RMWC[17]	0.83	0.56	0.71	0.54	0.65
MVC[11]	0.70	0.56	0.68	0.52	0.61
ObMiC[14]	0.87	0.68	0.82	0.71	0.77
Upper[	0.81	0.57	0.73	0.68	0.70
Ours-Expand	0.81	0.55	0.65	0.65	0.67
Ours Refine	0.85	0.64	0.73	0.68	0.72

TABLE II
PERFORMANCE OF IRRELEVANT FRAME DETECTION ON AND RELATED THE NEW VIDEO OBJECT COSEGMENTATION DATASET.

Category	All frames(I/R)	Error[18]	Our error(I/R)
Airplane	1763(61/1702)	20	39(33/6)
Balloon	1459(65/1394)	13	0(0/0)
Bear	1338(56/1282)	3	3(1/2)
Car	592(14/578)	4	8(5/3)
Eagle	1703(38/1665)	23	88(5/83)
Ferrari	1272(28/1244)	11	9(8/1)
Skating	1173(58/1115)	0	45(0/45)
Horse	1189(55/1134)	5	0(0/0)
Parachute	1461(40/1421)	14	125(12/113)
Single diving	1448(76/1372)	18	34(29/5)

B. New Video Object Co-segmentation Dataset

To further evaluate the performance of the proposed method for irrelevant frame detection, the proposed system also tests the method on the new video object co-segmentation dataset [18], which includes 101 videos and 10 categories of objects in total. And in particular, part of the videos contain irrelevant frames which greatly increase the difficulty. For each frame, a binary label is used to indicate the relevant/irrelevant frame. The number of misclassified frames is shown in Table II.

It can be noticed from Table II, although performed in an unsupervised way, the method still achieves competitive or even better results compared with the base line in 6 groups. For some other video groups, such as *eagle* and *parachute*, the error detections are mainly because that the targets are very ambiguous and no proper target object proposals are generated for them. However, the detection errors for irrelevant frame stay low for most of the videos, which show the effectiveness of the framework.

IV. CONCLUSION

A new video co-segmentation approach, which is based on common proposal clusters searching and adaptive recognition for irrelevant frames, is introduced in this letter. The common targets' mining strategy greatly improves the segmentation of each video, and the adaptive criterion condition adds flexibility for irrelative frames involved cases. The proposed framework is still extensible for multiclass problems. A straightforward extension could be conducting multiple selections for different common targets.

REFERENCES

[1] W. Brendel and S. Todorovic, "Video object segmentation by tracking regions," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep. 2009, pp. 833
[2] M. Grundmann, V. Kwatra, M. Han, and I. Essa, "Efficient hierarchi-cal graph-based video segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2141–2148.
[3] C. Xu, C. Xiong, and J. Corso, "Streaming hierarchical video segmenta-tion," in *Proc. Eur. Conf. Comput. Vis.*, 2012, vol. 7577, pp. 626.
[4] B. Alexe, T. Deselaers, and V. Ferrari, "What is an object?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 73–80.
[5] Y. J. Lee, J. Kim, and K. Grauman, "Key-segments for video object segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011,
[6] T. Ma and L. J. Latecki, "Maximum weight cliques with mutex con-straints for video object segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 670–677.
[7] D. Zhang, O. Javed, and M. Shah, "Video object segmentation through spatially accurate and temporally dense extraction of primary object regions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp628–635.
[8] X. Cao, F. Wang, B. Zhang, H. Fu, and C. Li, "Unsupervised pixel-level video foreground object segmentation via shortest path algorithm," *Neurocomputing*, vol. 172, pp. 235–243, 2016.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 12, July 2017

[9] D.-J. Chen, H.-T. Chen, and L.-W. Chang, "Video object cosegmenta-tion," in *Proc. 20th ACM Int. Conf. Multimedia*, 2012, pp. 805–808.
[10] J. C. Rubio, J. Serrat, and A. López, "Video co-segmentation," in *Proc. Asian Conf. Comput. Vis.*, 2012, pp. 13–24.
[11] W.-C. Chiu and M. Fritz, "Multi-class video co-segmentation with a gen-erative multi-video model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 321–328.
[12] J. Guo, Z. Li, L.-F. Cheong, and S. Z. Zhou, "Video co-segmentation for meaningful action extraction," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2232–2239.
[13] J. Guo, L.-F. Cheong, R. T. Tan, and S. Z. Zhou, "Consistent foreground co-segmentation," in *Proc. Asian Conf. Comput. Vis.*, 2014, pp. 241–257.
[14] H. Fu, D. Xu, B. Zhang, and S. Lin, "Object-based multiple fore-ground video co-segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3166–3173.
[15] Z. Lou and T. Gevers, "Extracting primary objects by video co-segmentation," *IEEE Trans. Multimedia*, vol. 16, no. 8, pp. 2110–2117, Dec. 2014.
[16] H. Fu, D. Xu, B. Zhang, S. Lin, and R. K. Ward, "Object-based multiple foreground video co-segmentation via multi-state selection graph," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3415–3424, Nov. 2015.
[17] D. Zhang, O. Javed, and M. Shah, "Video object co-segmentation by reg-ulated maximum weight cliques," in *Proc. Eur. Conf. Comput. Vis.*, 2014, vol. 8695, pp. 551–566.
[18] L. Wang, G. Hua, R. Sukthankar, J. Xue, and N. Zheng, "Video object dis-covery and co-segmentation with extremely weak supervision," in *Proc. Eur. Conf. Comput. Vis.*, 2014, vol. 8692, pp. 640–655.
[19] M. Bergtholdt, J. Kappes, and S. Schmidt, "A study of parts-based object class detection using complete graphs," *Int. J. Comput. Vis.*, vol. 87, no. 1–2, pp. 93–117, 2010.