Bag of Squares: A Reliable Model of Measuring Superpixel Similarity

Conference Paper · July 2014

DOI: 10.1109/ICME.2014.6890320

CITATIONS
3

READS
214

4 authors, including:

Shijie Zhang
Tianjin University
3 PUBLICATIONS **5** CITATIONS
SEE PROFILE

Wei Feng
Tianjin University
62 PUBLICATIONS **430** CITATIONS
SEE PROFILE

Jiawan Zhang
Tianjin University
101 PUBLICATIONS **404** CITATIONS
SEE PROFILE

BAG OF SQUARES: A RELIABLE MODEL OF MEASURING SUPERPIXEL SIMILARITY

Shijie Zhang[1,2], *Wei Feng*[1,2,*], *Jiawan Zhang*[3], *Chi-Man Pun*[4]

[1] School of Computer Science and Technology, Tianjin University, Tianjin, China
[2] Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin University, Tianjin, China
[3] School of Computer Software, Tianjin University, Tianjin, China
[4] Faculty of Science and Technology, University of Macau, Macau, China

ABSTRACT

As the increasing popularity of superpixel-based applications, measuring superpixel-level similarity becomes an important and commonly required problem. In this paper, we propose a general bag of squares (BoS) model for such particular purpose. Compared to existing methods, our approach provides a full scheme to both invariantly represent superpixels and accurately measure their pairwise similarities. In order to handle the split-and-merge variety of superpixels of same objects in different scenes, our model is based on superpixel pyramid. As a result, the BoS model of a superpixel is built upon a group of subregions consisting of the superpixel itself and its children subregions in the pyramid. For each subregion, we extract a proper number of maximum squares via distance transform, and then use a fast self-validated approach to clustering them into a small number of dominant squares, which together with a rotation and scale invariant square descriptor, jointly compose the BoS model for the particular superpixel. Finally, we measure the similarity between a pair of superpixels by the closeness of their BoS models. Experiments on interactive object segmentation and co-saliency detection show that the proposed BoS model can reliably capture the delicate differences among superpixels, thus always producing better segmentation results, especially for segmenting highly variant objects in clutter scenes.

Index Terms— Bag of Squares (BoS), superpixel-level similarity, scale and rotation invariance, image segmentation

1. INTRODUCTION

Superpixels [1] are perceptually meaningful irregular subregions of an image, which are usually generated by grouping neighboring pixels with similar appearances [2, 3]. Recently, many algorithms in computer vision and image processing use superpixels, instead of the original pixels, as the atomic primitives, e.g. object segmentation and detection [4, 5], tracking [6], and cosegmentation [7]. This is mainly because that using superpixels can significantly reduce the complexity of computation while maintaining comparable, sometimes even better, precision than using pixels [8].

One key problem in superpixel-level applications is how to reliably measure the similarity between two superpixels. However, most previous superpixel-related work mainly focuses on superpixel generation [2, 3] or its applications in different vision problems [1, 5, 7]. For similarity measuring, superpixels are simply treated as common image regions that are described using state-of-the-art regional features, such as regional histogram [4], region covariance [8], GMM

Fig. 1. BoS model of superpixels. (a)-(d) show the BoS representations for the SLIC superpixels [2] of two images with same foregrounds. In (a) and (c), the superpixels labeled as 1 and 2 correspond to the hair of two little girls, respectively. But, the same objects are separated into quite different number of superpixels with very different shapes and spatial connections in the two scenes. This is the split-and-merge variety property of superpixels. Superpixels labeled as 3 indicates the multiple appearances property of superpixels.

* is the corresponding author. Email: wfeng@tju.edu.cn. This work is supported by NSFC (61100121), the Program for New Century Excellent Talents in University (NCET-11-0365), the National Science and Technology Support Project (2013BAK01B01, 2013BAK01B05), and in part by Science and Technology Development Fund of Macau (008/2013/A1 and 034/2010/A2).

and SPM [7], and the similarity of a pair of superpixels are just measured as the closeness of their corresponding regional features. In spite of the reasonable performance of such method in some applications, it ignores the peculiarities of superpixels that may lead to the failure of common regional features in measuring superpixel-level similarities.

As shown in Fig. 1, compared to regular image regions, superpixels have the following three distinct properties: (1) irregularity, i.e. superpixels of an image may have very different sizes and shapes, (2) split-and-merge variety, i.e. the same object in two images may correspond to different number of superpixels with very different sizes, shapes and spatial connections, (3) multiple appearances, i.e. although superpixels are generally homogeneous subregions, some superpixels may have multiple homogeneous appearances. Although there are a number of successful regional image features, none of them are specifically designed for describing superpixels and catering the above three peculiarities.

In this paper, we propose a general bag of squares (BoS) model, which includes both BoS detector and BoS descriptor, to faithfully represent superpixels and accurately measure their pairwise similarities. Fig. 2 shows the algorithm flow of the proposed BoS model, which is based on superpixel pyramid that allows us to handle the split-and-merge variety of superpixels. Specifically, the BoS model of a superpixel is built upon a group of subregions that includes the superpixel itself as the root and its all children subregions in the superpixel pyramid. For each subregion, we extract a proper number of maximum squares via distance transform, and then use a fast self-validated approach to clustering them into a suitable but much smaller number of dominant squares. We use these dominant squares to represent a superpixel that enable us to capture multiple delicate homogeneous appearances in the superpixel. We then present a rotation and scale invariant square descriptor to finalize the BoS model. At last, the superpixel level similarity is measured as the closeness of corresponding BoS models. We have tested the performance of our BoS model on interactive object segmentation and co-saliency detection. Extensive results show that the proposed BoS model can reliably capture the delicate differences among superpixels, thus can always produce better accuracy, especially for highly variant objects in clutter scenes.

2. BAG OF SQUARES DETECTOR

As shown in Fig. 1, the BoS model regularly represents the irregular superpixels by variant numbers of dominant squares. In this section, we introduce the three major steps to efficiently extract such dominant squares that captures the multiple delicate appearances of superpixels and allows the usage of regular region descriptors.

Superpixel pyramid. As shown in Fig. 2, for a given image, we first generate an L-level superpixel pyramid by revising the SLIC algorithm [2].

Using a desired superpixel number K as input, SLIC initializes clusters by sampling K regularly spaced cluster centers and gradually moving them to locations with locally lowest gradient magnitude. Next, in the assignment step, each pixel i is associated with the nearest cluster center. Once all the pixels have been associated with the nearest cluster center, an update step is used to adjust the cluster centers to be the mean feature vector of all the pixels belonging to the cluster. Finally, a post-processing step is used to enforce connectivity by re-assigning disjoint pixels to nearby largest neighboring cluster. At the end of this process, we add a label mask M to enforce the same label covered pixels belonging to a same cluster. Within the pyramid, level $i+1$ contains more superpixels than level i, and all superpixels of level $i+1$ exactly obey the superpixel boundaries of level i.

To generate a superpixel pyramid, we first produce level 1 superpixel segmentation $\mathcal{P}_1$ by standard SLIC algorithm. Given superpixels segmentation $\mathcal{P}_i$ of level i as label mask M and the number of superpixels K_{i+1} ($K_{i+1} > K_i$), we generate the superpixels $\mathcal{P}_{i+1}$ of level $i+1$ by examining the belonging relationship of the initial regular seeds of $\mathcal{P}_{i+1}$ to the superpixel labels of $\mathcal{P}_i$. At level $i+1$, every initial seed may produce a superpixel containing only the pixels having the same label with the initial seed. Note, this method is able to generate superpixel pyramid via other regular seeds based superpixel algorithms, e.g. TurboPixel [3].

Regional maximum squares extraction. To handle the superpixel peculiarities, our BoS model is constructed for superpixels on the 1st level in the pyramid; and for each superpixel of the 1st level, we use all its children superpixels in the pyramid and itself as source subregions to extract a number of candidate maximum inscribed squares (or maximum squares for short). Our BoS model is based on the maximum squares of a subregion, since they are regular and maximally cover the appearance of the subregion. As shown in Fig. 2, for a subregion, we approximately extract its maximum square through the distance transform [9]. Since the distance map indicates the minimal distance from each pixel to the region boundary, we only need find the pixel with maximal distance value and use it as center and the distance value as radius to construct the maximum square. Note that, this solution is of linear complexity, thus is very fast. By subtracting the extracted maximum square, we can recursively extract a proper number of maximum squares from the given subregion.

Square main direction extraction. We use a simple moment-based method to extract the main direction of a superpixel square that is commonly used in feature detection [10]. For a particular square, we first compute its moment-based intensity centroid as

$$C = \big(M_{10}/M_{00}, M_{01}/M_{00}\big), \tag{1}$$

where M_{ij} is the intensity moment in the square defined as

$$M_{ij} = \sum_{x,y} x^i y^j I(x,y). \tag{2}$$

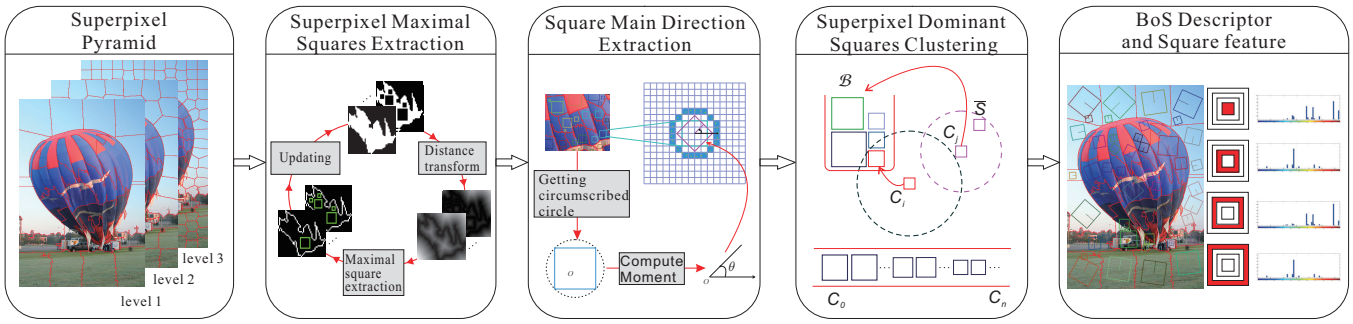

Fig. 2. Algorithm flow of the bag of squares (BoS) model and the BoS based superpixel-level similarity measurement.

From the square centroid C, we obtain its main direction as

$$\theta = \tan^{-1}(\frac{m_{01}}{m_{10}}). \tag{3}$$

In our implementation, all the moments in Eq. (2) are computed within the circumscribed circle region.

Dominant squares clustering. Now, for a superpixel P, we have a number of candidate maximum squares $\mathcal{C}_p$. As shown in Fig. 2, we sort $\mathcal{C}_P = \{C_0, \cdots, C_n\}$ in a descendant order according to their sizes. The dominant squares $\mathcal{S}_P = \{S_0, \cdots, S_n\}$ of superpixel P are selected from $\mathcal{C}_P$ by the following self-validated clustering process [11]. First, $S_0 = C_0$, i.e. the largest candidate square is a dominant square, and its corresponding weight w_0 is initialized as 1. Then, for the candidate square C_i in $\mathcal{C}_P$, we compare its similarities to all dominant squares in $\mathcal{S}_P$ with its similarity to the farthest sample of $\mathcal{S}_P$, denoted by $\bar{S}$, in the feature space. If C_i is closer to $\bar{S}$ than to any dominant squares in $\mathcal{S}_P$, we add C_i as a new dominant square into $\mathcal{S}_P$ with corresponding weight 1; otherwise, C_i is clustered into the closest square in current $\mathcal{S}_P$ by increasing the corresponding weight by 1.

As introduced in Sec. 3, we describe a single dominant square S using a sparse histogram. Thus, in the space of histograms, we can approximate the farthest sample of a dominant square S with sparse histogram as

$$\bar{S} = \frac{1 - S}{m - 1} \tag{4}$$

where m is the number of bins in the square histogram. Hence, in the histogram space, we can approximate the most impossible sample $\bar{S}$ of current dominant squares S_P as S_{mean}, where S_{mean} represents the mean sparse histogram of all dominant squares in S_P.

3. BAG OF SQUARES DESCRIPTOR

The BoS model of a superpixel P is a weighted set of dominant squares,

$$\mathcal{B}_P = \langle \mathcal{S}_P, W_P \rangle \tag{5}$$

where $\mathcal{S}_P$ is the set of dominant squares, and $W_P = [w_0, \cdots, w_{|\mathcal{S}_P|}]^T$is the weight vector denoting the importance of corresponding dominant squares. We now introduce an invariant descriptor for the BoS model, and based on which how to measure superpixel-level similarity.

Dominant square descriptor. As shown in Fig. 2, we further represent a single dominant square as a set of evenly distributed concentric bands, each of which is a regular region and can be accurately described using the quantized color histogram H that uses 16 bins to quantize each color channel, thus is a 4096 bins sparse histogram [4].

The dominant square descriptor is composed by sequentially stitching the histogram of concentric square bands from the inside out, which forms a larger sparse histogram after normalization. Finally, the BoS descriptor of $\mathcal{B}_P$ is defined as the set of sparse histograms of all dominant squares in $\mathcal{B}_P$.

BoS similarity measurement. For two squares S_u and S_v, we use the Bhattacharyya coefficient to measure their statistical closeness,

$$\rho(S_u, S_v) = \frac{1}{|c|} \sum_{i=1}^{|c|} Ba(H_u, H_v) \tag{6}$$

$$Ba(H_u, H_v) = \sum_{i=1}^{b} \sqrt{H_u(i) \cdot H_v(i)} \tag{7}$$

where H_u and H_v are histogram descriptors of square S_u and S_v, respectively, c is the number of square bands, and b is the number of histogram bins. Based on $\rho(\cdot)$, we can measure the unnormalized similarity of two BoS $\mathcal{B}_A$ and $\mathcal{B}_B$ as

$$\psi(\mathcal{B}_A, \mathcal{B}_B) = \frac{1}{Z} \sum_{S_u \in \mathcal{S}_A, S_v \in \mathcal{S}_B} w_u w_v \rho(S_u, S_v), \tag{8}$$

where w_u and w_v are the weights of squares S_u and S_v, respectively. $Z = \sum_{S_u \in \mathcal{S}_A, S_v \in \mathcal{S}_B} w_u w_v$ is the normalization factor.

Finally, for two superpixels P and Q, we measure their normalized similarity as the average pairwise similarity of all dominant squares in P and Q,

$$\text{Sim}(P, Q) = \frac{\psi(\mathcal{B}_P, \mathcal{B}_Q)}{\psi(\mathcal{B}_P, \mathcal{B}_P)\psi(\mathcal{B}_Q, \mathcal{B}_Q)} \tag{9}$$

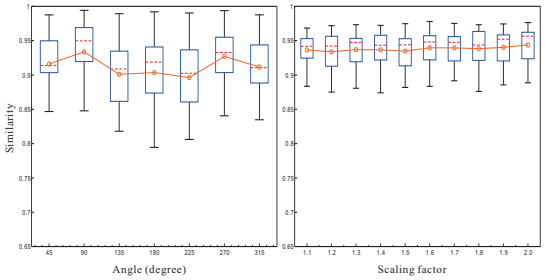

Fig. 3. Invariance of the BoS model to rotation and scaling.

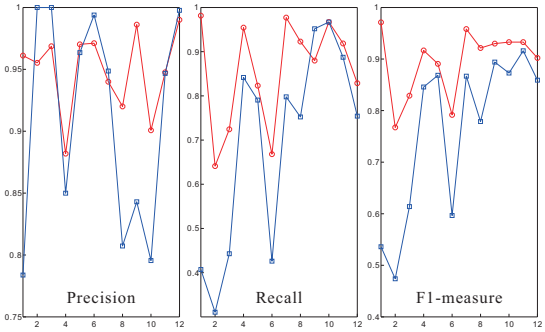

Fig. 4. Comparative segmentation precision and recall: BoS model (red) vs. superpixel regional histogram (blue).

where $\mathcal{B}_P$ and $\mathcal{B}_Q$ are the BoS of superpixel P and Q.

4. EXPERIMENTAL RESULTS

In this section, we verify the performance of the proposed BoS model on two superpixel-based applications, i.e. interactive image segmentation and image co-saliency detection.

Image segmentation. In our experiment, we evaluated the performance of the proposed BoS model in interactive foreground segmentation and compared with the state-of-the-art MSRM method [4].

We first tested the invariance of our BoS descriptor to image rotation and scaling. We randomly selected 5 images from the cosegmentation dataset [7, 8]. For each image, we generated 7 test rotation image by rotating the original image by the angles from $\frac{1}{4}\pi$ to $\frac{7}{4}\pi$ and 10 test scaling images, respectively. We measured the pairwise similarities of all test images to the original one. Fig. 3 shows that the proposed BoS descriptor is quite stable to rotation and scaling. This is because the dominant square and concentric square band representation are invariant to image rotation and scaling.

We then tested the accuracy of the BoS model in interactive image segmentation. Fig. 4 shows the comparative precision, recall, and F1-measure of using our BoS model and the superpixel regional histogram of the MSRM method [4] for segmenting 12 randomly selected benchmark images. We can clearly see that our BoS model outperforms the superpixel regional histogram representation, which is attributed to

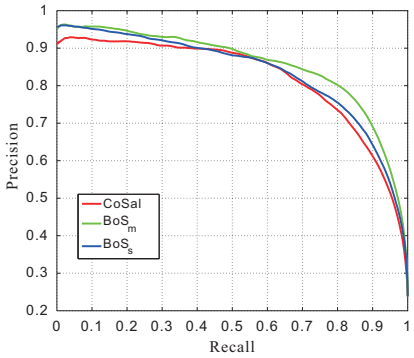

Fig. 6. PR curve of co-saliency detection on benchmark dataset: BoS model vs. histogram-based co-saliency detection [12]. BoS$_m$ represents the BoS model using multiple-layers superpixel pyramid and the BoS$_s$ denotes BoS using single-layer superpixels.

the abilities of BoS model to accurately capture the delicate appearance of superpixels and to better match the superpixel peculiarities than common regular region features. Fig. 5 shows more results for interactive foreground segmentation of using the BoS model and the MSRM method [4]. For the fairness of comparison, in each image, we used exactly the same (simple) foreground/background scribbles for both methods. The segmentation differences between MSRM and BoS are only caused by the superpixel representation and similarity measurement. Fig. 5 clearly shows that the superpixel regional histogram representation of MSRM may be able to generate reasonable results for homogeneous images, but may fail in segmenting highly textured and variant objects in clutter scenes, e.g. the leopard in forest or the doll in front of a tree. In contrast, our BoS model can produce much better results for such regions. Note that, in our experiments, we only used a small number of simple foreground/background scribbles, because in practice we usually prefer obtaining good segmentations using less number of scribbles.

Image co-saliency detection. In this experiment, we replace the regional color and texture feature descriptor of a state-of-the-art co-saliency model [12] with our BoS descriptor. That is, the color variation and texture property in each superpixel are described by our BoS model. Then we compare the performance of the modified co-saliency detection to the original one on the benchmark co-saliency dataset [12], which contains 105 image pairs covering multiple types of objects like human objects, flowers, buses, cars, and boats. Each image pair has ground truth masks about the human-labeled foreground and background. Fig. 6 shows the comparative PR curve of co-saliency detection using BoS descriptions, including multi-layered BoS (denoted by BoS$_m$) and single layered BoS (denoted by BoS$_s$), and the regional color and texture features. We can clearly observe that although the original method can achieve very good detection accuracy on the dataset, using BoS descriptions may further boost the

Fig. 5. Comparative results on interactive foreground segmentation.

performance. Besides, multi-layered superpixel pyramid produced the best and outperformed BoS_s that is better than the regional color and texture description.

Fig. 7 shows several co-saliency detection results using the BoS descriptors and compares with the results of original co-saliency model using regional color and texture feature [12]. The number in each co-saliency map indicates the F1-measure of this result. It can be seen that using BoS description can further contribute to improvement of detection accuracy both quantitatively and perceptually.

5. CONCLUSION

In this paper, we have proposed a general scheme, namely BoS, to faithfully represent superpixels and accurately measure their pairwise similarities. Compared to the existing regular region descriptors, the BoS model better matches the peculiarities of superpixels and is able to accurately capture the delicate appearances of superpixels and can reasonably handle the specific split-and-merge variety of superpixels. Experiments on interactive image segmentation and co-saliency detection demonstrate the superior performance of the BoS model in measuring superpixel-level similarities than existing methods, especially for the superpixels of highly variant objects in clutter backgrounds. In the future, we will be interested in investigating the application of the BoS model in multiple images cosegmentation and superpixel-based large displacement optic flow.

6. REFERENCES

[1] X. Ren and J. Malik, "Learning a classification model for segmentation," in *ICCV*, 2003.

[2] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE TPAMI*, vol. 34, no. 11, pp. 2274–2282, 2012.

Fig. 7. Comparative co-saliency detection results: BoS vs. CoSal [12]. The F_σ measure is showed in the corner of each image, that is obtained by the weighted mean of precision and recall[12] using $F_\sigma = \frac{(1+\sigma^2)\mathrm{Pre}\cdot\mathrm{Rec}}{\sigma^2\mathrm{Pre}+\mathrm{Rec}}$, where $\sigma^2 = 0.3$ following the convention of co-saliency detection.

[3] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, "Turbopixels: Fast superpixels using geometric flows," *IEEE TPAMI*, vol. 31, no. 12, pp. 2290–2297, 2009.

[4] J. F. Ning, L. Zhang, D. Zhang, and C. K. Wu, "Interactive image segmentation by maximal similarity based region merging," *Pattern Recognition*, vol. 43, no. 2, pp. 445–456, 2010.

[5] Y. Yang, S. Hallman, D. Ramanan, and C. Fowlkes, "Layered object detection for multi-class segmentation," in *CVPR*, 2010.

[6] S. Wang, H. C. Lu, F. Yang, and M. H. Yang, "Superpixel tracking," in *ICCV*, 2011.

[7] A. Joulin, F. Bach, and J. Ponce, "Multi-class cosegmentation," in *CVPR*, 2012.

[8] S. J. Zhang, W. Feng, L. Wan, J. W. Zhang, and J. M. Jiang, "Scalable image co-segmentation using color and covariance features," in *ICPR*, 2012.

[9] R. Fabbri, L. da F. Costa, J. C. Torelli, and O.M. Bruno, "2d euclidean distance transform algorithms: A comparative survey," *ACM Computing Surveys*, vol. 40, no. 1, pp. 1–44, 2008.

[10] P. L. Rosin, "Measuring corner properties," *Computer Vision and Image Understanding*, vol. 73, no. 2, pp. 291–307, 1999.

[11] W. Feng, J. Jia, and Z.Q. Liu, "Self-validated labeling of markov random fields for image segmentation," *IEEE TPAMI*, vol. 32, no. 10, pp. 1871–1887, 2010.

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