REALISTIC IMAGE COMPOSITE WITH BEST-BUDDY PRIOR OF NATURAL IMAGE PATCHES

Yuan Wang[1,2], *Fan Zhong*[1*], *Xiangyu Sun*[1], *Xueying Qin*[1,2]

[1]School of Computer Science and Technology, Shandong University

[2]Engineering Research Center of Digital Media Technology, Ministry of Education, P.R. China

ABSTRACT

Realistic image composite requires the appearance of foreground and background layers to be consistent. This is difficult to achieve because the foreground and the background may be taken from very different environments. This paper proposes a novel composite adjustment method that can harmonize appearance of different composite layers. We introduce the Best-Buddy Prior (BBP), which is a novel compact representations of the joint co-occurrence distribution of natural image patches. BBP can be learned from unlabelled images given only the unsupervised regional segmentation. The most-probable adjustment of foreground can be estimated efficiently in the BBP space as the shift vector to the local maximum of density function. Both qualitative and quantitative evaluations show that our method outperforms previous composite adjustment methods.

***Index Terms*—** Image Composite, Composite Adjustment, Appearance Transfer

1. INTRODUCTION

Image composite is a fundamental operation in image editing and processing. A typical scenario is to extract a foreground object from one image, and overlay with another background image. Foreground extraction has been deeply studied in the literature of computer graphics and computer vision [1, 2]. Given the foreground alpha matte, the overlay operation can be done easily with alpha blending. However, since the foreground and background are from different scenes, simply blending them together will result in unrealistic composite. To harmonize foreground and background appearances, photographers usually need to adjust standard image controls such as brightness,contrast,hue,and saturation [3, 4], which is tedious and unavailable for online applications.

Automatic composite adjustment has gained some attention in the past years. Lalonde et al. [5] learns color statistics of natural images, and apply it for natural image classification and realistic composite. The color statistics is represented as the color histogram of all available pixels. Xue et al. [6] identify key statistical measures of realism, then perform composite adjustment by selecting key zones in the global color histograms. These methods mainly use global statistics of foreground and background. However, due to the complexity of the background, the joint distribution of foreground and background global statistics are difficult to be modeled reliably. These methods thus often fail to discover the correct relations between foreground and background.

Regional and local image statistics has been investigated in some color transfer methods [7, 8]. In comparison with global statistics, local statistics can find better correspondences when source and target images have large differences [9]. However, image composite is different from color transfer in that no similar structure exists between foreground and background. To further improve correspondences, semantic information also has been exploited [10], which requires input images to be segmented and labelled, and thus applicable only for typical outdoor scenes.

In this paper we propose a novel composite adjustment method based on co-occurrence probability of local image patches, i.e. the *Best-Buddy Prior*(BBP). Instead of model the complete joint probability of all possible pairs of patches, BBP involves only those patch pairs from different regions that have high correlation in appearances and materials. Each such a patch pair is called a *Best-Buddy*, which can be discovered from unlabelled natural images, enabling BBP to be learned easily. In BBP space, the most probable foreground color shift can be estimated by searching the local maxima of density function.

2. BEST-BUDDY PRIOR

The best-buddy prior is used to measure the co-occurrence probability of image patches that have similar appearance and materials. Given an image I with region segmentation $\{R_k\}, \bigcup R_k = I$. $D(p, q)$ is a distance function measuring the similarity of two patches p, q. The *best-buddy* of $p, p \in R_i$ is defined as:

$$BB(p) = \arg min_q \ D(p, q), q \in \bar{R}_i$$

This work is supported by NSFC (No.61572290, 61672326), 863 program of China (No.2015AA016405).

which means that $BB(p)$ is nearest to p among all patches not in the same region with p.

We exclude patch pairs in the same region in order to exploit only co-occurrence probability between different regions/objects, otherwise $BB(p)$ may have high probability to be selected as the spatial neighbor of p. To produce the region segmentation $\{R_k\}$, we adopt the efficient graph-based segmentation method in [11], which is an unsupervised segmentation that does not need any interaction. Note that the region segmentation is used only to exclude patch pairs in the same region, so it is insensitive to under-segmentation, we thus use a large edge threshold in order to avoid over-segmentation.

The distance function $D(p, q)$ can be defined as requirements. Since our goal is to search patch pairs with similar appearance and materials, we adopt HOG [12] to represent the texture of patches, and define $D(p, q)$ as the distance of HOG descriptor and mean color, i.e. $D(p, q) = D_{hog}(p, q)+\lambda D_{color}(p, q)$. λ is a coefficient, and it is fixed to be 0.5 for all of our experiments. The HOG descriptor is computed with patch size 32×32. The color difference is computed in HSL space. Note that we use color as measure in order to avoid selecting patch pairs differ a lot in color, since HOG considers only intensity variations. Figure 1 shows examples of *Best-Buddy* computed by our program, each *Best-Buddy* is marked with rectangles in the same color.

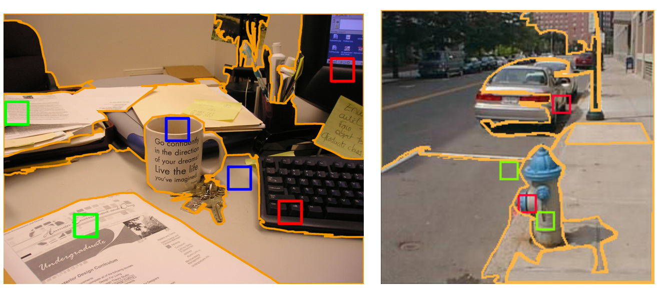

Fig. 1. Examples of best-buddies, rectangles with the same color are a pair of best-buddies.

2.1. Learning the Best-Buddy Prior

The best-buddy patch pairs can be extracted from any input image automatically. Given an image set $\mathcal{I} = \{I_i\}$, a set of best-buddies $\mathcal{B}$ can be extracted. Note that for each image I_i, a large number of distinct best-buddies can be extracted. However, since our goal is to model the distribution of best-buddies with respect to composite features, it is unnecessary to use all possible best-buddies. We thus randomly select a set of patches from each image, and for each patch p, we add $\{p, BB(p)\}$ into the set $\mathcal{B}$ if both p and $BB(p)$ are not across region boundary, so small regions and patches close to region boundaries will be prevent from being selected.

The best-buddy prior is used to model the distribution of $\mathcal{B}$ in a feature space $\mathcal{X}$ related with image composite. Let $\{p, q\} \in \mathcal{B}, q = BB(p)$ a best-buddy patch pair, $\mathcal{X}_{pq} = \{\mathcal{X}_p, \mathcal{X}_q\} \in \mathcal{X}$ is the corresponding feature descriptor composed with the descriptors of p and q. With these notations, the best-buddy prior can be represented as a probability density function $P(x), x \in \mathcal{X}$ modeling the distribution of best-buddies in natural images. Given $\mathcal{B}$ and $\mathcal{X}_{pq}$, $P(x)$ can be constructed easily with Kernel Density Estimation (KDE).

We compute the descriptors $\mathcal{X}_{pq}$ using the averages of patch statistical measures. It has been proved that luminance L ,saturation S and correlated color temperature(CCT) C have a strong correlation in background and foreground [6]. Therefore, these measures are involved in descriptors. We use Zone System [13, 14, 6] which is well developed in photography to divide histogram for luminance. Specifically, we divide statistical measure histogram into high, middle, and low zones. We only use high zone of luminance and saturation for computing the descriptors because the mean of high zone matches better than the mean of whole histogram [6]. $\mathcal{X}_{pq}$ is then defined as follows:

$$\mathcal{X}_{pq} = \{L_p, C_p, S_p, L_q, C_q, S_q\} \tag{1}$$

where $L_p, C_p, S_p, L_q, C_q, S_q$ are the means of luminance, CCT, saturation of patches p, q.

To learn the prior probability $P(x)$, we collect an natural image set containing 8000 images with different illumination and scene categories. Note that these images does not need any labelling, so the size of training image set is easy to be extended. From these images we extract 141036 best-buddies, and use the corresponding feature descriptors to estimate the density function $P(x)$.

$P(x)$ measures the occurrence probability of best-buddy patch pairs in natural images. In other words, larger $P(x)$ means more realistic and less violation of perceptual naturalness for the corresponding regions. For example, a light foreground is less likely to occur in a dark background, for such a composite image, the best-buddies connecting foreground and background will have low probabilities with respect to $P(x)$.

3. COMPOSITE ADJUSTMENT BASED ON BEST-BUDDY PRIOR

Composite adjustment aims to adjust the foreground appearance in order to be matched with the background. Following the description in section 2.1, we adjust the foreground by shifting its mean luminance L, saturation S, and CCT C, which requires to estimate the foreground shift vector $\mathcal{V} = \{\Delta L, \Delta C, \Delta S\}$.

If a composite is unrealistic, a best-buddy $\{f, b\}$ between foreground and background will have low prior probability $P(\mathcal{X}_{fb})$. The most-probable composite that looks realistic, should be indicated by a point $\tilde{\mathcal{X}}_{fb}$ that is close to $\mathcal{X}_{fb}$, and at the same time have high prior probability. We determine $\tilde{\mathcal{X}}_{fb}$ by using mean-shift method, with $\tilde{\mathcal{X}}_{fb}$ be the local maxima point searched start from $\mathcal{X}_{fb}$. Note that in this way we avoid using hard threshold to filter low probability positions. The shift vector produced from $\mathcal{X}_{fb}$ is then computed as:

$$\mathcal{V}(\mathcal{X}_{fb}) = (\tilde{\mathcal{X}_f} - \tilde{\mathcal{X}_b}) - (\mathcal{X}_f - \mathcal{X}_b) \tag{2}$$

which is equivalent to:

$$(\mathcal{X}_f + \mathcal{V}(\mathcal{X}_{fb})) - \mathcal{X}_b = (\tilde{\mathcal{X}_f} - \tilde{\mathcal{X}_b})$$

so adding this shift vector to the composite foreground can equalize the difference between foreground and background to the training feature. This is similar to the mean-matching approach in previous color transfer methods [15].

For each composite, we randomly select K best-buddies $\mathcal{X}_{fb}^k, k = 1, \cdots, K$ between foreground and background. The final shift vector is computed as:

$$\mathcal{V} = \sum_{k=0}^{K} \omega_k \mathcal{V}(\mathcal{X}_{fb}^k) \tag{3}$$

with ω_k a weighting function defined as follows:

$$\omega_k = \frac{1}{Z} e^{-\frac{(x_f - x_b)^2 + (y_f - y_b)^2}{\sigma}} \tag{4}$$

in which Z is a normalization factor, (x_f, y_f), (x_b, y_b) are the locations of foreground and background patches in the composite, respectively.

4. EXPERIMENTS

In this section, we experimentally evaluate the proposed composite adjustment method and do comparisons with previous methods. Since the quality of composite is difficult to be measured, we propose to do quantitative evaluation with a set of synthetic unrealistic composite images.

To generate the synthetic dataset with groundtruth, we utilize 60 real image from the *LabelMe* data-set [16]. We adjust objects' appearances(lighting, white balance and saturation) in those real images randomly, so that the composite images looks unrealistic. All of those objects and background are from natural images, therefore they are semantically reasonable and the groundtruth are realistic. We take the adjusted unrealistic composite images and foreground mask as inputs of different composite adjustment methods. For the proposed approach, we use the luminance channel(l) in decorrelated $l\alpha\beta$ space [17, 15], because the change made in luminance channel affect values in other channels a little. Saturation is the saturation channel of HSV space. We calculate CCT using the package OptProp [18]. Figure 2 shows the prodedures and intermediate results of our method.

Figure 3 shows four examples of the adjusted results using five methods: simple cut-and-pase (the input unrealistic composite), Photoshop Match Color, the method of Lalonde and Efros [5](labeled as ColorComp), the method of Xue Su et al. [6](labeled as ZoneComp), our method, and groundtruth. We compute the results of ColorComp and ZoneComp using codes provided by the authors. Our method performs well even when background and foreground are very different, and produces results that are more close to the groundtruth compared with other methods[1].

To quantitatively evaluate the results, we compute the mean absolute error(MAE) of results with the ground truth. MAE is computed as:

$$MAE = \frac{1}{n} \sum_{i=0}^{n} \mid (v_{result}(i) - v_{groundtruth}(i)) \mid \tag{5}$$

where n is the number of pixels in the foreground region, v(i) is the value of pixel i in RGB color space. The lower error indicates that the result is more similar to the ground truth.

Fig.3 shows the quantitative comparisons of our method with previous approaches. Note that for better visualization, the MAE value is normalized with the MAE of input composite. We can find that for most examples, our method achieves better results than previous methods. Talbe 1 compares average errors over all test images, our method achieves significantly lower error rates.

Our method can achieves better results because BBP is a more accurate prior than the priors used in the previous methods. The method in [6] and Photoshop matchColor(using the method in [15]) are based on mean-matching approach, which assumes that the offsets between foreground and background are close to zero. This assumption may be invalid in some situations and is difficult to model complicated interactions between foreground and background. On the other hand, note that the adjustment method in section 3 always determine foreground shift vector based on matched local maxima in BBP space, so it is less likely to be influenced by noise and errors in constructing the best-buddy prior (segmentation and matching error, etc.). This is also why we can search best-buddies with simple HOG features and colors, without semantic analysis and accurate matching of materials.

5. CONCLUSION

In this paper we introduce Best-Buddy Prior, and apply it for automatic composite adjustment. BBP is a novel method to model the interactions between natural image regions, and can be learned efficiently with unlabelled images. The proposed composite adjustment method is evaluated quantitatively with synthetic unrealistic test images, results show that our method can produce more realistic composite in comparison with previous approaches. BBP has potential to be used for other applications, such as realism assessment, which we will consider in future works.

[1]The results of all the 60 test images are included in the supplemental materials.

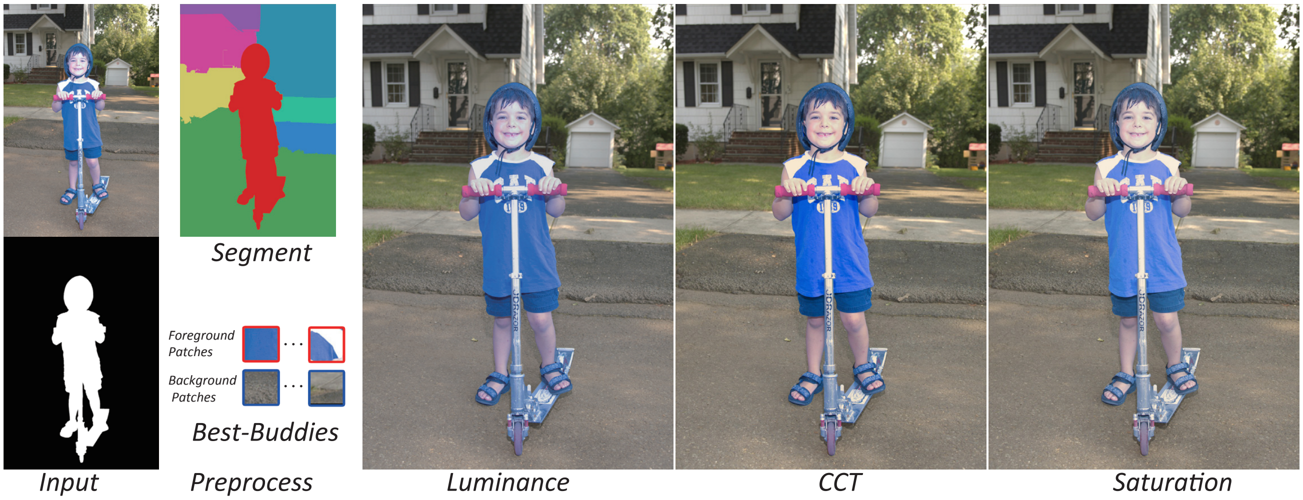

Fig. 2. The procedures and intermediate results of the proposed adjustment method. The right three images are the results sequentially adjusted with luminance, CCT, saturation.

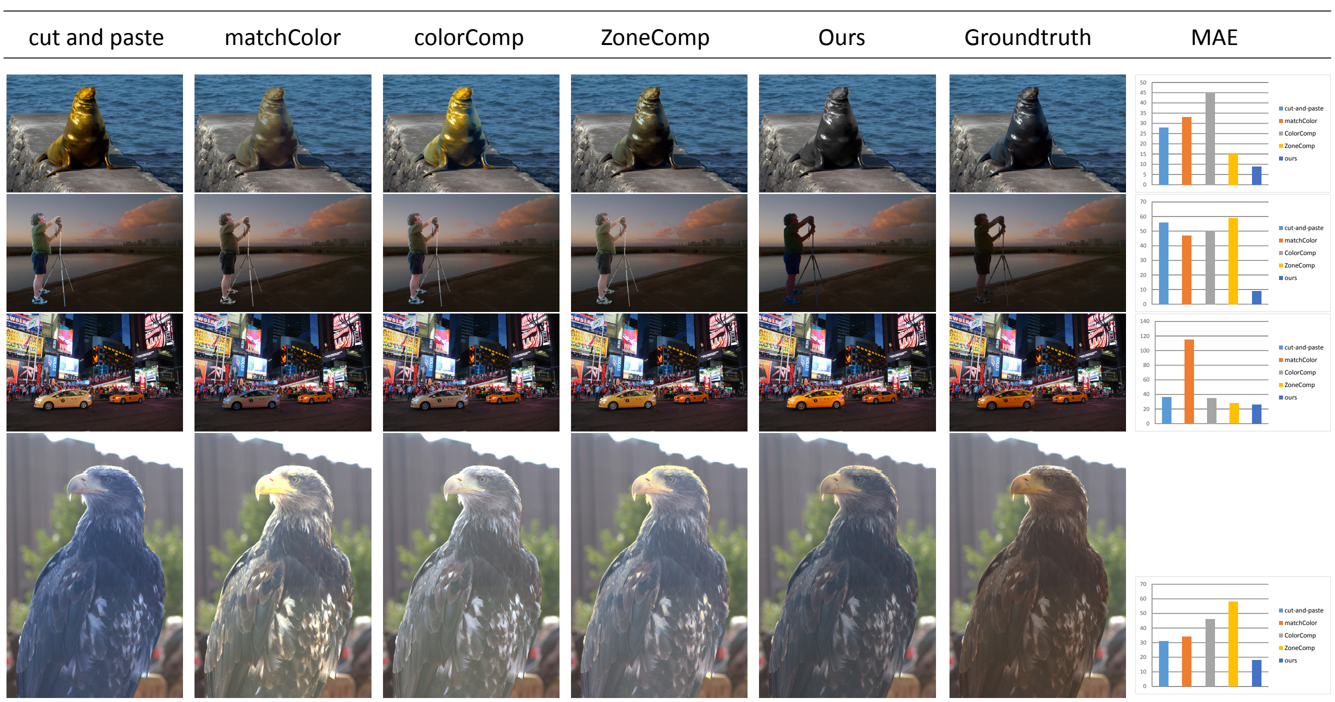

Fig. 3. Comparisons of our method with cut-paste, MatchColor,ColorComp, and ZoneComp. The last column show MAE of different methods.

	MatchColor	ColorComp	ZoneComp	our method
average($\frac{MAE_{result}}{MAE_{input}}$)	1.0332	1.1532	1.0433	0.7874

Table 1. Average MAE errors of different methods on the synthetic dataset.

6. REFERENCES

[1] Alvy Ray Smith and James F Blinn, "Blue screen matting," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996, pp. 259–268.

[2] Christoph Rhemann, Carsten Rother, Jue Wang, Margrit Gelautz, Pushmeet Kohli, and Pamela Rott, "A perceptually motivated online benchmark for image matting," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 1826–1833.

[3] C Damerval, S Meignen, and V Perrier, "A fast algorithm for bidimensional emd," *Signal Processing Letters IEEE*, vol. 12, no. 10, pp. 701–704, 2005.

[4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[5] Jean-Francois Lalonde and Alexei A Efros, "Using color compatibility for assessing image realism," in *IEEE International Conference on Computer Vision*. IEEE, 2007, pp. 1–8.

[6] Su Xue, Aseem Agarwala, Julie Dorsey, and Holly Rushmeier, "Understanding and improving the realism of image composites," *Acm Transactions on Graphics*, vol. 31, no. 4, pp. 13–15, 2012.

[7] Yu-Wing Tai, Jiaya Jia, and Chi-Keung Tang, "Local color transfer via probabilistic segmentation by expectation-maximization," in *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*. IEEE, 2005, vol. 1, pp. 747–754.

[8] Weiming Dong, Guanbo Bao, Xiaopeng Zhang, and Jean-Claude Paul, "Fast local color transfer via dominant colors mapping," in *ACM SIGGRAPH ASIA 2010 Sketches*. ACM, 2010, p. 46.

[9] Fuzhang Wu, Weiming Dong, Yan Kong, Xing Mei, Jean-Claude Paul, and Xiaopeng Zhang, "Content-based colour transfer," *Computer Graphics Forum*, vol. 32, no. 1, pp. 190–203, 2013.

[10] Yi-Hsuan Tsai, Xiaohui Shen, Zhe Lin, Kalyan Sunkavalli, and Ming-Hsuan Yang, "Sky is not the limit: Semantic-aware sky replacement," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, pp. 149, 2016.

[11] Pedro F. Felzenszwalb and Daniel P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.

[12] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, pp. 886–893.

[13] *The Ansel Adams photography series*, Little, Brown and Company.

[14] Lu Yuan and Jian Sun, "Automatic exposure correction of consumer photographs," *Lecture Notes in Computer Science*, vol. 7575, no. 1, pp. 771–785, 2012.

[15] Erik Reinhard, Michael Ashikhmin, Bruce Gooch, and Peter Shirley, "Color transfer between images," *IEEE Computer Graphics & Applications*, vol. 21, no. 5, pp. 34–41, 2001.

[16] By Bryan C. Russell, Bryan C. Russell, Antonio Torralba, Antonio Torralba, Kevin P. Murphy, Kevin P. Murphy, William T. Freeman, and William T. Freeman, "Labelme: A database and web-based tool for image annotation," in *MIT AI Lab Memo*, 2010.

[17] Daniel L Ruderman, Thomas W Cronin, and Chuan-Chin Chiao, "Statistics of cone responses to natural images: Implications for visual coding," *JOSA A*, vol. 15, no. 8, pp. 2036–2045, 1998.

[18] J Wagberg, "Optprop, matlab toolbox for calculation of color related optical properties," 2007.