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

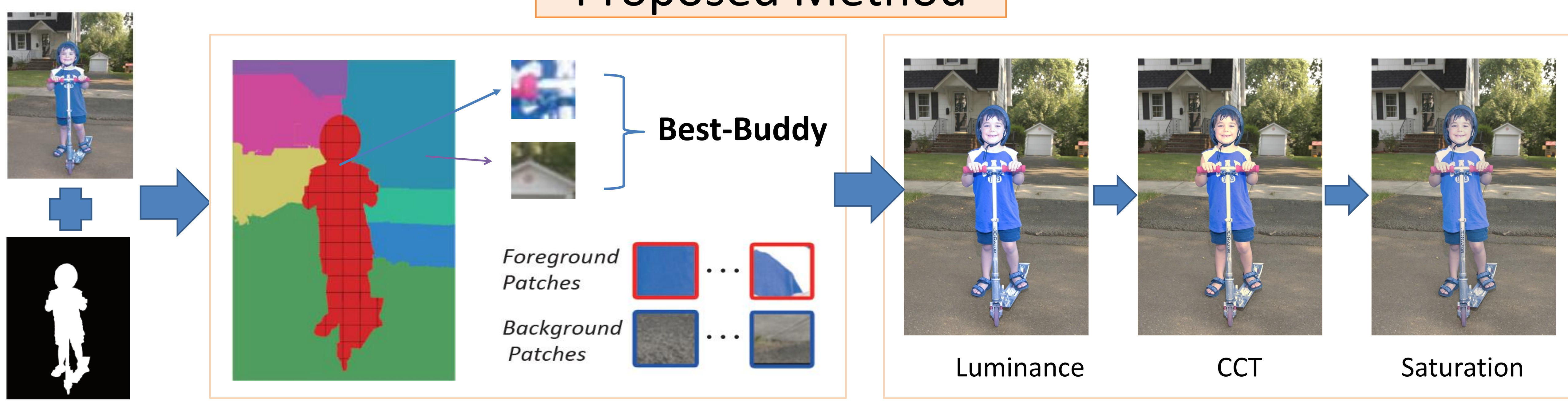
Yuan Wang[1,2], Fan Zhong[1], Xiangyu Sun[1], Xueying Qin[1,2]
[1]School of Computer Science and Technology, Shandong University
[2]Engineering Research Center of Digital Media Technology, Ministry of Education, P.R. China

Problem Statement

Background Object Cut-Paste Result Realistic Image

Combining background and object, we get a composite image. However, to get a realistic image composite is very difficult. 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 Best-Buddy Prior (BBP), which is a novel compact representations of the joint co-occurrence distribution of natural image patches.

Proposed Method



input preprocess Composite adjustment

We adjust the foreground by shifting its mean luminance ***L***, saturation ***S***, and CCT ***C***, which requires to estimate the foreground shift vector $V = \{\Delta L, \Delta S, \Delta C\}$ If a composite is unrealistic, a best-buddy 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 $\widetilde{\chi_{fb}}$ that is close to $\mathcal{X}_{fb}$, and at the same time have high prior probability. We determine $\widetilde{\chi_{fb}}$ by using mean-shift method, with $\widetilde{\chi_{fb}}$ 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:

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

For each composite, we randomly select *K* best-buddies X_{fb}^k between foreground and background. The final shift vector is computed as:

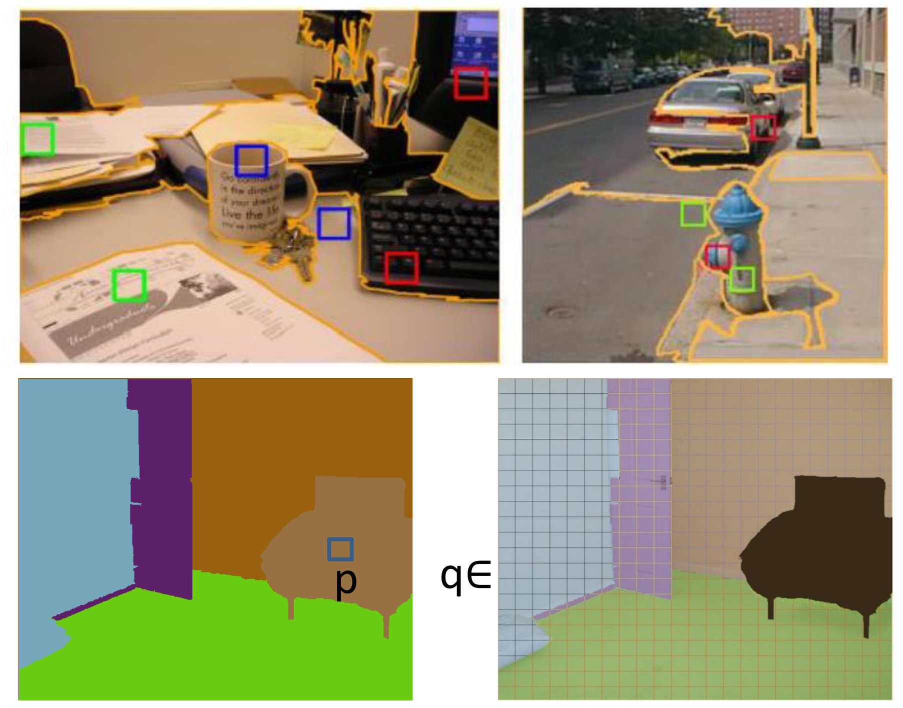
$$V = \sum_{k=0}^{K} \omega_k V(X_{fb}^k)$$

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

Conclusion

In this paper we introduce Best-Buddy Prior, and show its application for automatic image composite. Our method adopt *Best-Buddy* to estimate the foreground shift and outperform previous approaches. First, we estimate PDF of *Best-Buddies*. Then, we use prior to compute the local maximum for each *Best-Buddy* in composite. Finally, we shift histogram using the information in shift vector. For future work, it's of great interest to explore how to find similar material patch using more efficient technology.

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:

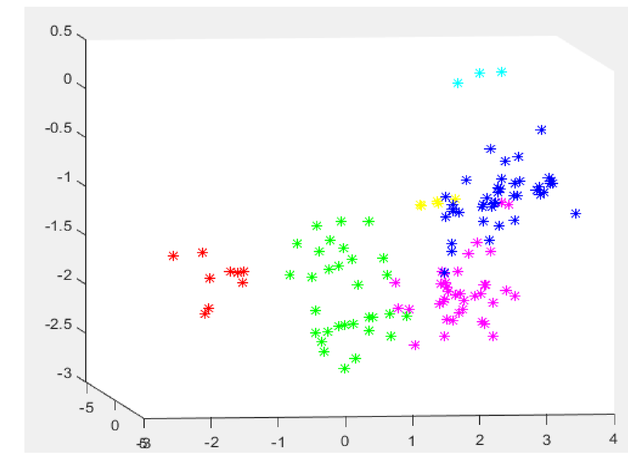
$$BB(p) = \arg min_q \; D(p,q), q \in \bar{R}_i$$

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

The best-buddy prior is used to model the distribution of $\mathcal{B}$ in a feature space $\mathcal{X}$ related with image composite. Let $\{p,q\} \in \mathcal{B}$, $q=BB(p)$ is 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). $\mathcal{X}_{pq}$ is then defined as follows:

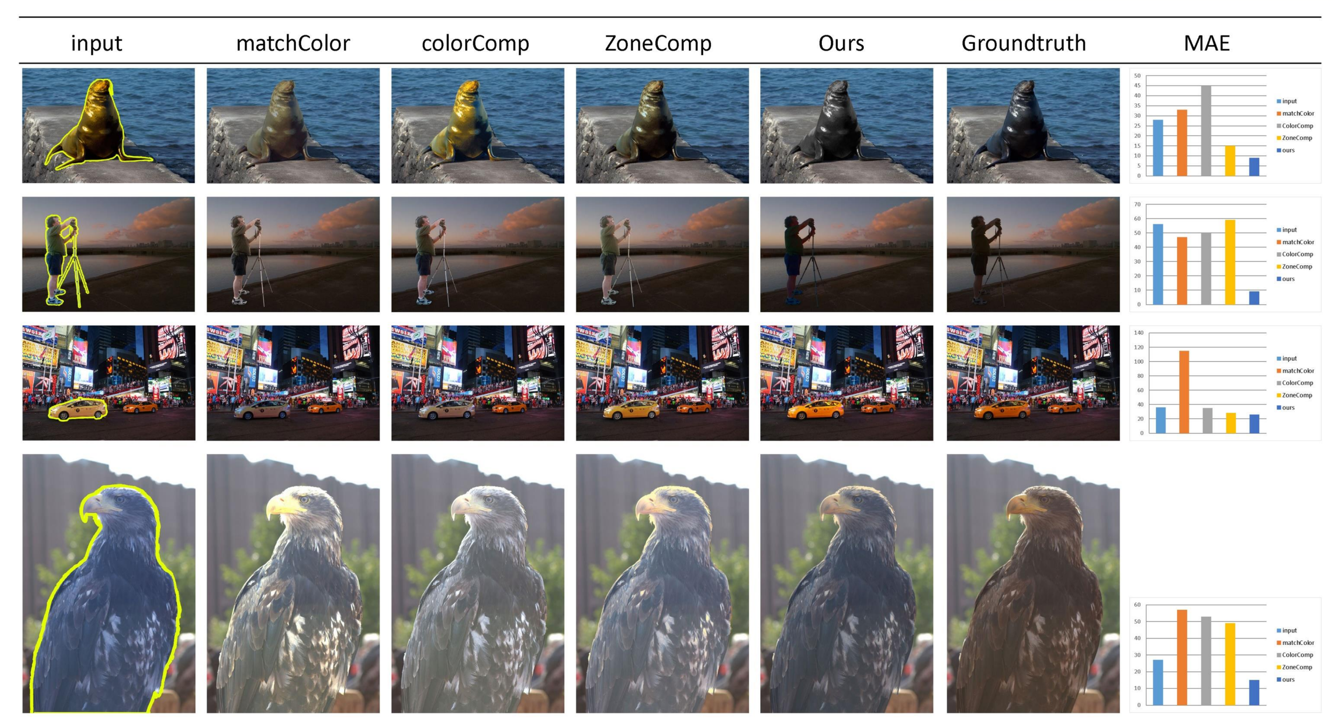
$$\mathcal{X}_{pq} = \{L_p, C_p, S_p, L_q, C_q, S_q\}$$

where L_p, C_p, S_p are the means of *luminance, CCT, saturation* of patches *p* , *q*. *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).*

Results



	MatchColor	ColorComp	ZoneComp	our method
average($\frac{MAE_{result}}{MAE_{input}}$)	1.0501	1.1612	1.0424	0.7871

Figure 3

$$MAE = \frac{1}{n} \sum_{i=0}^{n} |\,(v_{result}(i) - v_{groundtruth}(i))\,|$$

Dataset

To generate the synthetic dataset with groundtruth, we utilize 60 real image from the LabelMe data-set [16]. We adjust objects' appearances (saturation, white balance, and lighting) in those real images randomly, so that the composite images looks unrealistic.

Comparison

Figure3 shows four examples of adjusted results using four methods: Photoshop Match Color, the method of Lalonde and Efros [5](labeled as ColorComp), the method of Xue Su etal. [6](labeled as ZoneComp), our method, and groundtruth.
To quantitatively evaluate the results, we compute the mean absolute error(MAE) of results with the ground truth.

About Me

Wang Yuan (Email: wyelaine92@qq.com)

I'm looking forward to an opportunity to study for a PhD degree.

- I got my MSc in Computer Science from Shandong University(China) 2017.
- I has been a research intern in Chinese Academy of Sciences(CAS) from 2016.

Interests：

- Image Processing
- Style/color transfer
- computer vision for visual arts