DOI: 10.2352/ISSN.2470-1173.2016.14.IPMVA-374

Data-driven Approach to Aesthetic Enhancement

***Jihye Choi* [a]*, Sungjoon Koh* [a]*, Jongwoo Kwack* [a]*, Yonghun Kwon* [a] *and Hyunjung Shim* [a]**
[a] *School of Integrated Technology and Yonsei Institute of Convergence Technology,*
Yonsei University, Seoul, South Korea

Abstract

Traditional image enhancement techniques revise the distribution of pixels or local structure and achieve the impressive performance in image denoising, contrast enhancement and color adjustment. However, they are not effective to improve the overall aesthetic image quality because it may involve contextual modifications, including the removal of disturbing objects, inclusion of appealing visual elements or relocation of the target object.

In this paper, we propose a new aesthetic enhancement technique that edits the structural image element guided by a large collection of good exemplars. More specifically, we remove/insert image elements and resize/relocate objects based on good exemplars. Additionally, we remove undesirable regions determined by user interaction and fill these holes seamlessly guided by the exemplars. Based on the experimental evaluation on the database of two landmarks, we observe the considerable improvement in aesthetic quality.

Keywords: Aesthetic Enhancement, Data-driven approach, Image Transfer Technique

1 Introduction

Influenced by smart phones and wearable devices with built-in cameras, more pictures than ever are taken and shared nowadays. Unlike professional photographers, ordinary users do not always take aesthetically good pictures; they often lack the skills in finding good scene configuration, view or illumination. For example, imagine the pictures of the Eiffel tower taken at dusk before it light up, the Hollywood Sign slightly taken out of the frame or the Empire State Building obscured by other buildings in New York. Some other times, your pictures may be ruined by undesirable subjects. For instance, garbage trucks suddenly disturb the perfect street photo or other tourists accidently appear in the landmark photo. In order to assist ordinary users from these discouraging moments, we develop the data-driven approach to aesthetic enhancement by editing the object-level configuration of image.

Various image editing methods have been developed in the past and they adapt the variant of color filters [1] or re-colorization based on the intrinsic image decomposition [2]. Although they are effective in some applications such as image denoising, contrast enhancement and color adjustment, these image enhancement approaches are not effective to achieve an overall aesthetic enhancement (Figure 1). This is because the aesthetic enhancement may involve contextual modification in the mid-level segment (i.e. an object), rather than the low-level primitive (i.e. a pixel). For instance, in the case of Eiffel tower pictured at dusk, applying enhancement filters do not make the Eiffel tower more distinguishable. In this case, enhancing the target object that dominates the overall quality of the image in the principle of art would be more effective. Hence, we focus on the mid-level segment of input and modify the contextual information of input. They include the object relocation, illumination changes or removals of disturbing objects. In this way, we create appealing photographs so to achieve the aesthetic enhancement.

Original image　Filter in Instagram [1]　Proposed method

Figure 1. *Results from the image filtering and our object-level approach.*

In this paper, our goal is to develop the technique for the aesthetic enhancement. For that, we propose a new image transfer algorithm that implants appealing subjects directly from exemplar images. First, we construct the database of good exemplars. Given an input image, we search the best K exemplars which appearance matches well with the input. Once the top K exemplars are selected, we use them to fill the undesirable regions removed by user interaction and insert the appealing objects from exemplars. After that, we relocate/resize the objects to be fitted into the input.

When accessing the aesthetic quality of image, it is impractical to construct a closed form formulation that encapsulates all the variations such as color or illumination distribution, scene configuration or human perception. Hence, our transfer strategy is guided by a massive database of eminent images collected from online social communities. In recent years, data-driven approaches have been actively studied in scene completion [3], depth transfer [4] and super-resolution enlargement [5] and show that leveraging a huge image database leads to the big quantitative leap in performance. Inspired by the previous success, we construct the database, the large collection of good exemplar images, which well represents the complex variations of our solution space. Then, this database serves our model of appealing photographs.

This paper is organized as follows. In the following section, we introduce our algorithm; 1) the data collection and feature selection scheme, and 2) the transfer technique that consists of target removal, region completion and target repositioning. Then, experimental results are demonstrated and analyzed in section 4. Finally, we conclude the paper by highlighting the effectiveness of proposed method in section 5.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.14.IPMVA-374

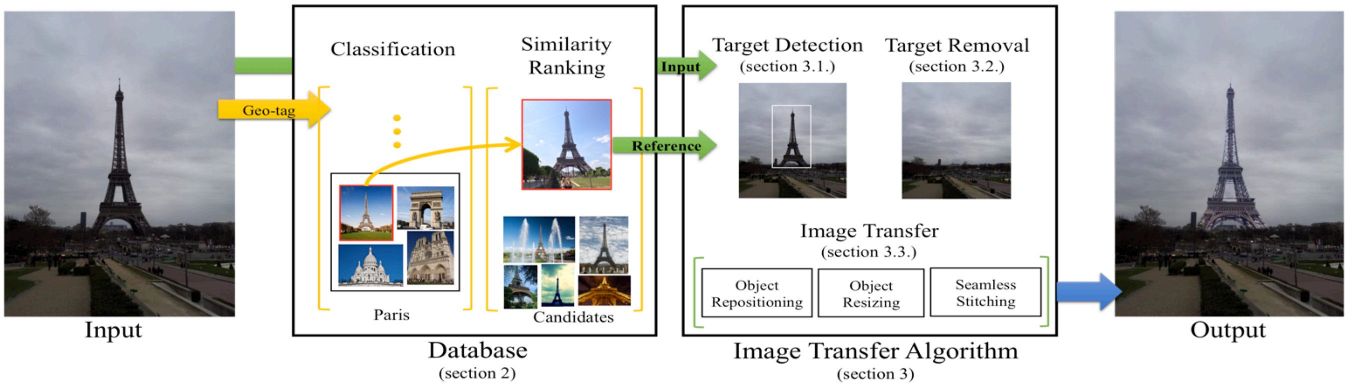

***Figure 2.** Overview of the proposed algorithm.*

2 Database and Image Retrieval

2.1. Database construction

To keep the efficient search and retrieval of image, our image database is confined by popular landmarks in the world. For the database construction, we fix several categories predetermined by their geographical regions and download eminent photographs from Flickr under keywords such as “Eiffel tower” or “Statue of Liberty”. We discard duplicate images, as well as the images whose horizontal dimension is below 1024 pixels. We keep aside low resolution ordinary photos to be used as the input.

Per exemplar image, the target object is defined by the tourist attraction itself and solely segmented from the background by Graph-cut algorithm [6] (Figure 3). This is to prepare segmented subjects that will be transferred into the input in section 4.

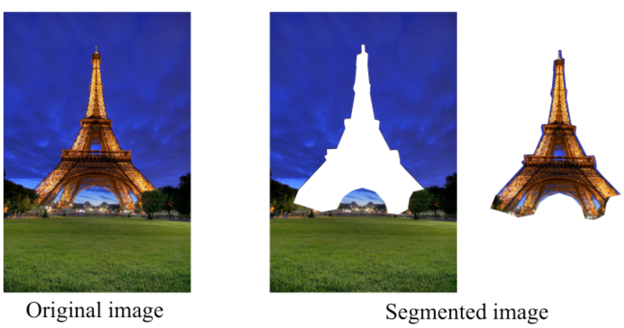

***Figure 3.** An example pair of eminent photograph and its segmented version. Our target is the construction itself, and it is masked out. We keep the original image, another version where the missing region is excluded and the segmented subject.*

2.2. Image retrieval from the database

Searching the best exemplar is important because the visual similarity between the exemplar and input influences the quality of image transfer. Given the input as in Figure 2, we use a geo-tag data (e.g. GPS) to classify its category. Within the category, we further retrieve the K most similar images to the input and they are chosen as the references. To define the similarity between images, we employ the HOG (Histogram of Oriented Gradients) descriptor [3]. We augment the HOG descriptors with RGB colormap to examine the color distribution between the images, as well as the composition when retrieving the image. More specifically, we first compute the feature descriptor and downsample the extracted features from the size of original input into 256 x 256. Then, we rank the images of the category based on the Euclidean distance in the feature space.

The top K (=five) images are demonstrated to the user and the user can select the preferred one among five ranked images as a reference image (Figure 4). If the user interaction is not available, our top ranked image can be used as the final reference image. Because the top ranked images are likely to share photographic characteristics (e.g., a view point, color distribution and background compositions) with the input, their image elements can be composited into the input without much visual artifacts.

Admittedly, the ranked images are not always in the same order with what humans heuristically ranked. Nonetheless, we could find out that it always contains the acceptable images.

3 Image Transfer Algorithm

We have defined the target construction when building the database. Therefore, once the reference photograph is selected, the template of target is available. Then, we localize the target from the input and exchange by the alternative in the reference. Our transfer algorithm consists of four steps as follows: 1) roughly identifying the position of target in the input, 2) erasing the target and completing the removed region, 3) transferring the alternative from the reference and 4) smoothing the abrupt boundaries of the inserted object.

3.1. Target detection

First of all, we determine the target object to be replaced. In the landmark photographs, the tourist attraction can be the most significant factor and thereby we consider it as the target object of our transfer algorithm. Hence, the first step is to identify the position of target in the input. We use the object mask as a template, encoding the segmented region in the reference. Then, we search the target from the input using the hierarchical template matching [6]. Note that our template encodes not only the color distribution but also the silhouette of target object. In fact, we find that this silhouette provides the robust clue for template matching. Although the viewing condition and color distribution are different between the template and input, the silhouette retains the structural

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.14.IPMVA-374

Figure 4. *Input image and similar images retrieved from the database. In this paper, we concentrate on the images of "Eiffel Tower" leaving other landmark architectures out of consideration. Top five images are displayed and the user chose the reference image.*

similarities between the template and target. To handle variable sizes of target objects, we resize the template within the range of maximum and minimum size and evaluate the matching cost in the sum of squared distance sense. The following pseudo-code describes our implementation on the hierarchical template matching.

Table 1: The pseudo-code of hierarchical template matching with input and template images.

```
Function HIERARCHICAL_TEMPLATE(Image input,
            Image template)
    [max_ratio, min_ratio] = Get_Size_Range(Image
                        input, Image template)
    While (min_ ratio ≤ r ≤ max_ ratio)
      resized_template = Resize Template given r
      s = Template_matching(Image input, Image
          resized_template)
      Insert s into array matching_score
    End
    Find_max_value(Array matching_score)
End function
```

In Table 1, we define the function 'Get_Size_Range', which determines the maximum and minimum size of template to compare with the input and template. We set the maximum size by the size of input image and the minimum rate by one quarter of the input size. Then, we apply the template matching algorithm varying the patch size among four levels between the maximum and minimum rates. The result of hierarchical template matching is illustrated in Figure 5.

3.2. Target Removal

Given the position of target object, a Graphcut algorithm [7] is applied to further refine the boundary of target object. By removing all pixels inside the boundary of target object, we successfully erase the target from the input, resulting holes. Then, we adopt the image inpainting algorithm [8] to fill these holes and the filled image is used as a background before the insertion of the template. Since the performance of image completion depends on the patch size, we apply with different patch sizes and the corresponding results are illustrated in Figure 6. In general, the small patch size tends to introduce the local variation into the filled region while the large patch size enforces to keep the overall variation. In our application, the small patch leads to the implausible inpainting results. Based on the empirical study, we choose the patch size to be greater than or equal to 71 x 71.

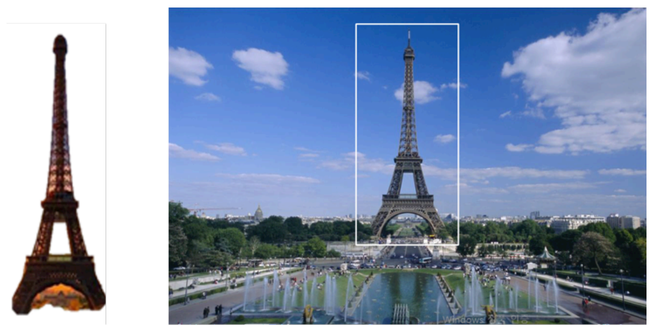

Figure 5. *The result of hierarchical template matching. It localizes the block that contains the target structure.*

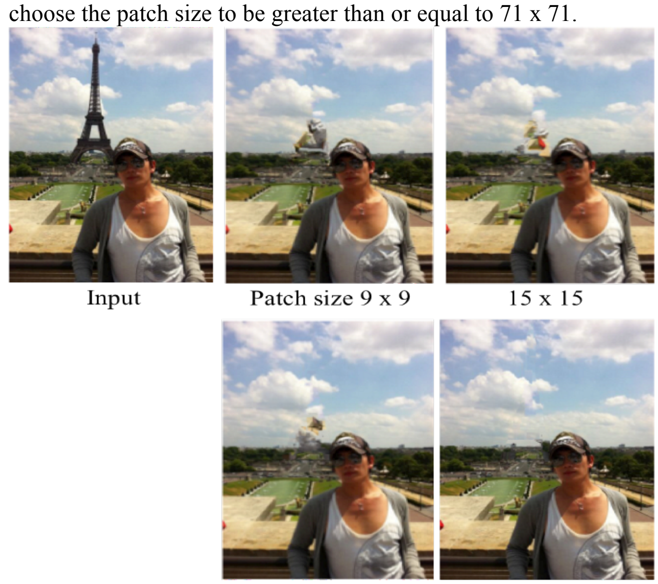

Figure 6. *Background images created by the inpainting algorithm with different patch sizes and their original input.*

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.14.IPMVA-374

3.3. Image transfer

From the image retrieval step, we ensure that the reference is visually similar to the input, meaning the viewing condition, illumination distribution and structural configuration. Hence, a simple insertion of template also makes a reasonable sense. However, our aim is to improve the aesthetic quality so that we push the transfer algorithm further to refine the size and position of target object. For that, our image transfer process is followed by object repositioning, resizing and seamless stitching. This fine-tune makes the output more appealing to humans.

When transferring the segmented target from the reference to the input, we set the bottom line for positioning the object. The bottom line is the bottommost pixels at the target object and the object is aligned along this line. Instead of the bottom line, it is possible to match the centroid of template and target. However, this can result the positioned target to be at a different composition from the input photograph (Figure 7); the Eiffel tower located in the sky or the tower of Pisa abruptly obscuring the subjects in the background such as the tourist, trees or cars. Based on the various case studies using different methods, we find that aligning the template to the bottom line yields more reliable results. After repositioning the template, we resize the inserted template. We measure the distance between the top and the bottom pixels in the target and resize the inserted object to fit into the target size. This process guarantees that the inserted target is placed inside the frame.

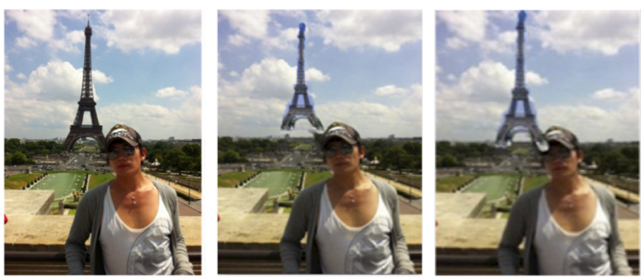

Figure 7. *The results of different object positioning methods.*

As shown in the third column of Figure 7, the boundaries between the background and the inserted template change very abruptly; introducing the artifacts in the photograph. To alleviate the unpleasant seams along the object boundaries, we apply a gradient editing technique [9] to minimize the visible artifacts.

4 Experimental Results and Discussion

We perform the processes described in section 2 and 3 sequentially to obtain our transfer results. We test the proposed algorithm on Eiffel tower images recorded at different viewpoint or time slot. Figure 8 demonstrates the examples of our results. In general, our results improve the details of target object and the object configuration within the scene. In the second and fourth row of Figure 8, the Eiffel tower appears slightly slanted due to the camera tilt. By applying our transfer method, the tower stands straight up. The quality of our results is dependent on the performance of image retrieval, i.e., how well it retrieves the exemplar images from the database. Obviously, the exemplar with the similar appearance to the input generates the appealing output. From Figure 8, the examples in third and fourth rows show that the selected reference is from the nearly same viewpoint of input. Moreover, since the color distribution between the input and the reference is similar, the ordinary human hardly recognizes pixels from the template. In fact, our results from the first and fourth row present the background from reference, i.e., the trapezoid-shaped area at the center of the Eiffel tower. Because both the input and exemplar have similar color distribution, it is possible to compensate the mismatch of background. On the contrary, there are some cases when retrieving good exemplar is not enough to achieve the aesthetic enhancement. As the third result in Figure 8 shows, even though the input and reference are pictured at same camera angle, the output is almost same with the reference. It is not our goal to transform images because we want to preserve the scene configuration of the original photograph, but these events occur when our target is the single object in the image and the background is monotonous.

When compositing the template into the input, both hole filling and gradient blending play important roles in reducing the seams along the object boundaries. In the fourth row, the background after the hole filling absorbs the textures from the surrounding ground and this degrades the quality of background. This adds some unreasonable textures at the bottom of Eiffel tower of our final result.

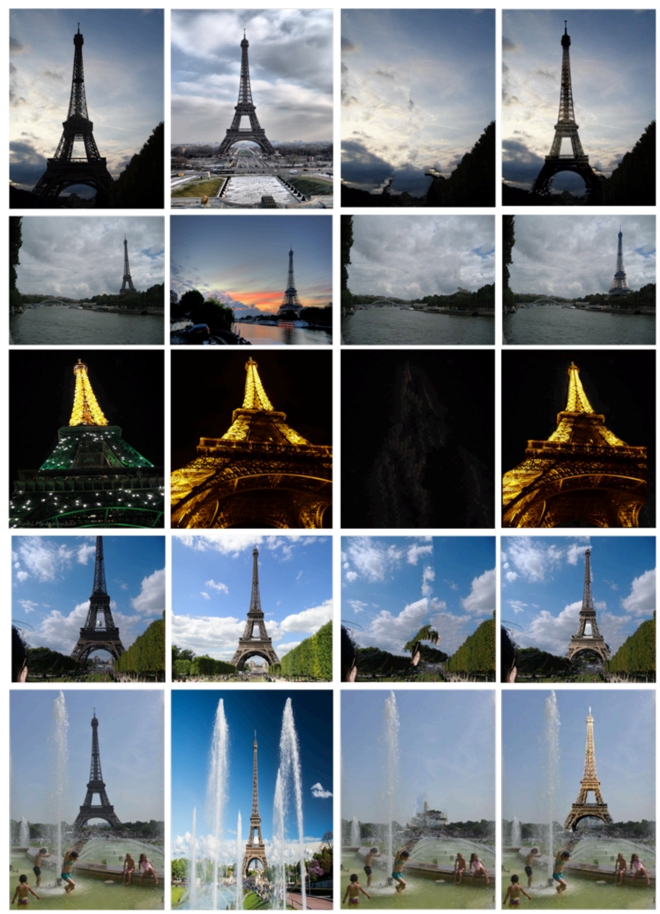

Figure 8. *Example results of our approach to aesthetic enhancement. Input and reference images are composited together to create the output image..*

DOI: 10.2352/ISSN.2470-1173.2016.14.IPMVA-374

5 Conclusion

We have proposed an image transfer approach to aesthetic enhancement that involves the contextual modification in object level. Our algorithm is guided by a large collection of images. We retrieve the reference as similar as possible to the input in terms of visual similarity; scene configuration and color or illumination distribution. Considering the reference as a good exemplar, we transfer the detail and configuration of an object into the input. As a result, we enhance the aesthetic quality of an ordinary photograph by modifying the objects that cannot be enhanced by applying pixel-level filters. Our algorithm has advantages that it modifies objects in image while preserving the scene configuration to be faithful to the input image.

Acknowledgement

This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the "IT Consilience Creative Program" (IITP-2015-R0346-15-1008) supervised by the IITP(Institute for Information & Communications Technology Promotion) and also the Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT and Future Planning (NRF-2013R1A1A1062644).

References

[1] https://www.instagram.com/

[2] Grosse, R., Johnson, M. K., Adelson, E. H., & Freeman, W. T. (2009, September). Ground truth dataset and baseline evaluations for intrinsic image algorithms. In Computer Vision, 2009 IEEE 12th International Conference on (pp. 2335-2342). IEEE.

[3] Hays, J., & Efros, A. A. (2007). Scene completion using millions of photographs. ACM Transactions on Graphics (TOG), 26(3), 4.

[4] Karsch, K., Liu, C., & Kang, S. B. (2014). Depth Transfer: Depth Extraction from Video Using Non-Parametric Sampling. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 36(11), 2144-2158.

[5] Freeman, W. T., Jones, T. R., & Pasztor, E. C. (2002). Example-based super-resolution. Computer Graphics and Applications, IEEE, 22(2), 56-65

[6] http://docs.opencv.org/doc/tutorials/imgproc/histograms/template_matching/template_matching. html

[7] Yuri Boykov, Vladimir Kolmogorov, "An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision". IEEE Trans. Pattern Anal. Mach. Intell. 26(9): 1124- 1137 (2004)

[8] Criminisi, A., Pérez, P., & Toyama, K. (2004). Region filling and object removal by exemplar-based image inpainting. Image Processing, IEEE Transactions on, 13(9), 1200-1212

[9] http://graphics.cs.cmu.edu/courses/15-463/2014_fall/hw/proj3g-gradient/