Optik 125 (2014) 4985–4989

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Multi-option image completion based on semantic matching image

Hao Wu*, Zhenjiang Miao

School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, PR China

ARTICLE INFO

Article history:
Received 28 August 2013
Accepted 5 April 2014

Keywords:
Image composition
Semantic matching
Global context
Local context

ABSTRACT

In this paper, we propose a new image completion method based on high-accuracy semantic matching. Existing image completion approaches mainly focus on simple filling, ignoring creativity and accuracy in semantic matching. Our method can complete missing regions with more creative and semantically matching images as well as create a more seamless and consistent completed image.

We use global and local features to search matching images for the target image. We then complete the missing region using Poisson blending and blending optimization. Results of experiments on challenging image databases show that the images are greatly improved. Thus, these results demonstrate the superiority of the proposed algorithm over existing image completion approaches.

1. Introduction

Image completion (also called hole-filing or inpainting) is used to replace or fill in the missing region of an image with new data in such a way that the modification cannot be detected. Previous research details two classical strategies for image completion.

In the first strategy, the missing region is reconstructed by using other images as accurately as possible. In recent years, studies have completed missing pixels using data that has already been there. The most significant methods involve extending adjacent contours and textures into the missing region [1–5]. This concept uses example-based texture synthesis [6–9] and additional constrains to preserve Gestalt cues explicitly, such as good continuation [10], either manually [11] or automatically [1]. None of the image completion approaches used in this strategy are creative.

In the alternative strategy, a plausible method is applied to complete the missing region using data that has already been there. The accurate reconstruction of the alternative method makes use of other sources of data to fill in the image, such as multiple photographs of the physical scene [12,9] or video [13].

This section places particular emphasis on [14], which used the gist model to classify images. The gist model can provide additional global information from the images under normal circumstances, but at the same time prevents high-accuracy image matching. By contrast, completing the missing region with content from other images allows more novel objects and textures to be inserted into the image.

Currently, high-performance image classification [15–19] and scene understanding [20–25] still need to extract local and global information from images. Certain classical algorithms discussed in these studies contribute to the present study.

Our study adopts a new method to complete missing regions using data from a massive number of images. This decision is based on two points, as follows. (1) The unknown region can be completed using only image data from the source image. (2) Even if the source image contains valid content, obtaining more new, creative images is difficult.

However, certain challenges are posed by using other creative images from the database. The first challenge is the difficulty of searching for semantically valid images. The second challenge is the difficulty of obtaining a consistent and seamless image composition. This paper attempts to solve these challenges.

In summary, our contributions include:

(1) Combining the local context matching and global context matching to achieve more accurate semantic matching.
(2) Completing seamless and consistent image composition.

2. Image matching based on semantic similarity

In this step, to complete an image, we search for semantically matching images by combining global and local contexts. We briefly introduce Gist and SIFT models. We then use context-occurring probability histograms to calculate the similarity of the two images and to search for semantically matching images.

* Corresponding author.
E-mail address: 10112056@bjtu.edu.cn (H. Wu).

http://dx.doi.org/10.1016/j.ijleo.2014.04.005
0030-4026/

2.1. Introduction of Gist

Gist [14], the global characteristics information of a scene, can be described by the low-dimensional signature vector of a scene. Gist descriptors are widely used in scene classification and recognition. The gist of a scene can represent the global context, which includes low-level features to high-level information. Gist can be learned on both conceptual and perceptual levels.

In this study, the Gist descriptor gathers the oriented edge responses at multi-scale into very coarse spatial bins. The Gist descriptor used in our study is built from six oriented edge responses at five scales gathered onto a 4×4 spatial resolution, thereby providing maximum effectiveness. We make use of the color information of the query image to augment the scene descriptor.

We calculate the SSD (I, R_u) (the Euclidean distance of the Gist descriptor between two regions), where R_u is the candidate regions labeled from the image database and I is the original image with missing region R_u. Lastly, we consider SSD (I, R_u) as an important reference of the semantic matching of the images.

2.2. Introduction of SIFT

The SIFT descriptor is used to extract the features of the invariant images. These features are invariant and highly distinctive to image rotation and scale. Semantic similarity also can be computed based on SIFT descriptor.

To extract the SIFT descriptors of the image, we must achieve the following four steps: (1) scale-space extrema detection, (2) keypoint localization, (3) orientation assignment, and (4) keypoint descriptor.

2.3. Context-occurring probability histogram

A total of 110 local semantic contexts are manually defined (car, desk, house, tree, and so on) to provide an abundant semantic content for image description.

For each semantic local context, we determine a classifier using a support vector machine (SVM) with kernel. For local contexts, the classifiers are determined using descriptions of randomly sampled image regions (motorcar, bike, people, bird, animal). After the annotation, approximately 300 relevant images are used for each context. These images are regarded as positive images for the connected context, whereas images from other contexts are considered as negatives. After obtaining the classifier of the local scene, we calculate the context-occurring probability of the local contexts.

2.4. Context-occurring probability

In this step, although we do not use the SVM classifier to directly classify the images, we use it to calculate the context-occurring probability (Maximum likelihood). We define the local contexts $c_1, c_2, \ldots, c_n$ (car, tree, desk, sky) and then calculate the occurring probability of these local contexts using the SVM classifier.

$p(c_1|k)$ can be computed using the outputs of the local context classifiers on all image regions in image k. These probabilities are calculated by running the connected context classifiers on the whole image. Finally, an image k is represented by the occurrence probabilities of local contexts: $(p(c_1|k), p(c_2|k), \ldots, p(c_n|k)$.

We choose one typical image as an example to introduce the context-occurring probability histogram.

Figs. 1 and 2 show that the context-occurring probabilities of apparent local contexts in the image are remarkably high (people, tree, cloud, house, car, chair). The context-occurring probability histogram can reflect the local information in the images.

Fig. 1. Typical image with a few local contexts.

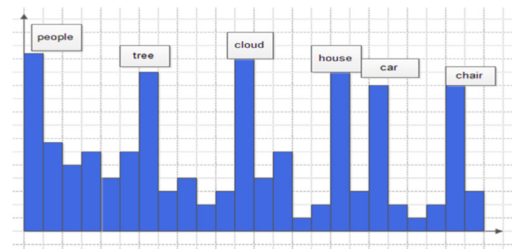

Fig. 2. Context-occurring probability histogram based on Fig. 1.

2.5. Similarity of histograms

Upon obtaining the context-occurring probability histograms, we use these histograms, as well as the classical distance, to calculate the semantic similarity of two images. This step is given by Eq. (1):

$$D(l) = \sqrt{1 - \sum_i \frac{\sqrt{H_1(i)H_2(i)}}{\sum_i H_1(i) \sum_i H_2(i)}} \tag{1}$$

In Eq. (1), $H_1(i)$ and $H_2(i)$ are the context-occurring probability histograms of the two images.

2.6. Combination of similar representations

Up to this point, we have constructed two types of image semantic similarity representations (global similarity representations $D(g)$ and local similarity representations $D(l)$). To combine the two representations, we use Eq. (2):

$$D = C \times D(g) + D(l) \tag{2}$$

We can adjust the parameter c to change the weight of either $D(g)$ or $D(l)$ if necessary. We then calculate $D(c)$ from the original image and the other images in the database. Finally, we use $D(c)$ to select semantically matching images from the database.

3. Massive images composition

After obtaining the semantically matching regions, we composite each matching region onto the incomplete image with the best possible placement by using graph cut seam finding [26] and Poisson blending [27]. Poisson editing is used extensively in computer vision; we only give a brief introduction of this method. The Poisson equation can achieve seamless filling of the boundary conditions.

However, after several groups of experiments, the resulting image composition is not realistic, particularly if the color dynamics of the incomplete image and that of its matching image are very different. To make sure that the composited image has consistent color, we use color handling [28] to process the color prior to composition.