PERFORMANCE IMPROVEMENT OF GENERIC OBJECT RECOGNITION BY USING SEAM CARVING AND SALIENCY MAP

Motoaki SATO and Jiro KATTO

Dept. of Computer Science, Graduate School of Fundamental Science and Engineering
Waseda University
3-4-1 Okubo, Shinjuku-ku, Tokyo, JAPAN
E-mail: {satoh, katto}@katto.comm.waseda.ac.jp

ABSTRACT

Recently, a demand for object recognition is increasing due to explosive increase of digital images. In the field of the general object recognition, we need many learning images whose content is already-known. However, it is difficult to construct such a learning dataset beforehand by ourselves. Therefore, it is general to use learning datasets that are made by third parties. However, many images of the datasets contain wide background areas other than the object area which is focused in the recognition. We think that this is one of the factors which bring recognition difficulty, and we aim at removing areas that are thought to be unnecessary beforehand to improve learning efficiency and recognition rate of the object recognition.

Keywords: Generic object recognition, Image classification, SIFT, Bag-of-features, Seam Carving, Saliency Map

1. INTRODUCTION

The general object recognition is a technology that makes a computer recognize a name of the object included in an image. Recently, this technology has advanced rapidly by the advance of computer vision and machine learning technologies. In this paper, we carry out experiments by using the bag-of-features method (Bof method) proposed by G.Csurka as a technique of image recognition [1]. In our experiments, we try to remove the areas that seem to be unnecessary for recognition by applying Seam Carving [3] and Saliency Map [4] to both the learning images and the test images. Seam Carving is an image resizing technique which iteratively removes pixel sets called "seam" having low energies in an image. Saliency Map is a visual model which provides "salient" regions in an image which are expected to be gazed by human eyes.

2. RELATED WORK

2.1 Bag of Features [1]

Bag-of-features representation is an analogy of bag-of-words representation for text categorization.
In bag-of-words, sentences are summarized to be sets of words by disregarding the word order. In bag-of-features, an image is represented by sets of local features (keypoints) by disregarding the position. Bag-of-features is also called Bag-of-keypoints.

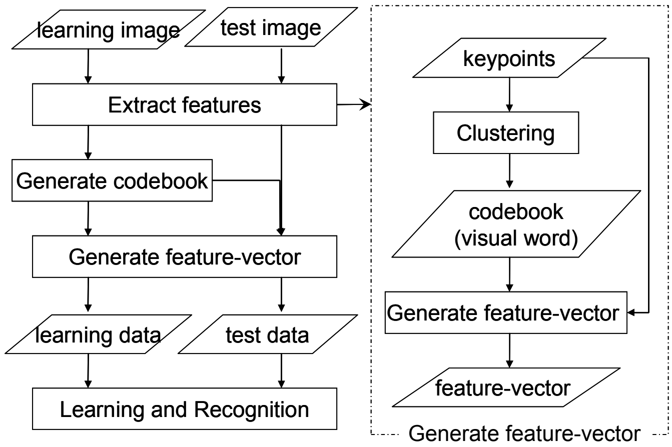

Fig. 1: Flowchart of bag-of-features.

Figure1 shows a general flowchart of the bag-of features method:

1) Extract feature points from images. Sparse sampling, dense sampling, etc. are used for the feature point extraction. Sparse sampling is a method to detect feature points by using Difference of Gaussian that is used in SIFT. Dense sampling is a method to set grid points or random points as feature points.

2) Describe local descriptors (keypoints) for extracted feature points. SIFT descriptor is a general technique to describe local features.

3) Execute clustering for vector quantization. A codebook is created by the clustering of keypoints. Centers of gravity of each cluster are regarded as representative vectors. They are called "visual words".

4) Generate feature vectors by using the codebook (visual words). In bag-of-features, feature vectors are expressed by the histogram of the number of visual words.

5) Learn or recognize an object by using the feature vectors. Feature vectors generated from the learning images are learning data, and feature vectors generated from the test images are test data.

2.2 SIFT [2]

SIFT (Scale-invariant feature transform) is a technique to detect and describe local features based on the appearance of an object at particular interest points. SIFT features are invariant to image scale and rotation. They are also robust to changes in illumination, noises and small viewpoint changes.

SIFT consists of two stages; detection of keypoints and description of features. The detection stage consists of scale space extreme detection and localization of key points. The description stage consists of assignment of orientation and extraction of SIFT descriptors. By the above steps, SIFT descriptors achieve robustness that is invariant to illumination (by the normalization of SIFT descriptors), rotation (by the assignment of orientation) and scale (by the scale space extreme detection).

2.3 Seam Carving [3]

Seam Carving supports content-aware image resizing for both reduction and expansion. A seam is an optimal 8-connected path of pixels on a single image from top to bottom, or left to right, where optimality is defined by an image energy function. By repeatedly carving out or inserting seams in one direction, we can change the aspect ratio of an image. By applying these operators in both directions, we can change the image size to a target size.

Energy values are defined for all pixels by the energy function. Therefore, Seam-carving can be considered as a problem of deciding the seam of which total energy is minimized. Next is a simple energy function:

$$e_1(I) = \left|\frac{\partial}{\partial x} I\right| + \left|\frac{\partial}{\partial y} I\right| \qquad (1)$$

where I is an $n \times m$ image. Define a vertical seam by

$$s^x = \{s_i^x\}_{i=1}^n = \{(x(i), i)\}_{i=1}^n, s.t. \forall i, |x(i) - x(i-1)| \le 1 \qquad (2)$$

Then, a seam of which energy is minimized is chosen according to the next equation, where function e is the energy value of seam s in image I.

$$s^* = \min_s E(s) = \min_s \sum_{i=1}^n e(I(s_i)) \qquad (3)$$

Figure 2 shows an example of horizontal and vertical seams, which is shown in the original paper [3].

Fig. 2: An example of horizontal/vertical seams [3].

2.4 Saliency Map [4]

Saliency Map is a vision model, that expresses which area attracts visual attention when an image is shown. It is expressed by three or more feature maps (simple visual features) at the early stage of human's visual information processing. Afterwards, those feature maps are integrated and a saliency map is generated.

3. PROPOSAL

3.1 Usage of SIFT, Seam Carving and Saliency map

In this section, how to apply SIFT, Seam-carving and Saliency Map to the proposal is explained. Figure 3 is an example image for a reference.

Fig. 3: The original image.

3.1.1 SIFT

SIFT uses sparse sampling or dense sampling for the feature point extraction. Sparse sampling is a method to detect feature points by using Difference of Gaussian. Dense sampling is a method to set grid points or random points to be feature points. In this paper, we use a dense sampling to set grid point as feature points. Therefore, only the description stage of SIFT is used. Figure 4 shows an example of SIFT descriptors based on dense sampling.

Fig. 4: SIFT-descriptors applied to Figure 3.

3.1.2 Seam Carving

Seam-carving is a technology that reduces an image size by removing seams of which energies are the minimum in an image. We apply next steps in our proposal. Function E is an energy of a seam in an image, and it is generally represented by an amount of gradient of pixels.

$$s^* = \min_s E(s) = \min_s \sum_{i=1}^{n} e(I(s_i)) < k \qquad (4)$$

At this time, we removed the area that seemed to be unnecessary for recognition by setting the threshold k in the energy value. In this paper, the value of k was set to 500. Figure 5 shows an example of Seam Carving applied to Figure 3. In Figure 6, SIFT descriptors are furthermore applied to Figure 5.

Fig. 5: A result of Seam-carving applied to Figure 3.

Fig. 6: SIFT-descriptors applied to Figure 5.

3.1.3 Saliency Map

Seam-carving has a problem of deciding a seam of which total energy is minimized without considering human attention (i.e. saliency). Besides, all pixels have energy values defined by a uniform energy function over an image. It can be thought that the saliency becomes higher in a foreground object region than in a background region. Therefore, areas that seem to be necessary for recognition can be enhanced by introducing Saliency Map to the energy map calculation. Concretely speaking, the method of applying Saliency Map is to allocate a luminance value of Saliency Map to an energy value of each pixel.
Figure 7 shows an example image in which Saliency Map is applied to energy map calculation in Seam Carving. Furthermore, Figure 8 demonstrates SIFT descriptors applied to Figure 7. In Figure 8, SIFT descriptors on grid points which are not salient are suppressed because the description areas of keypoints are limited to highly salient areas.

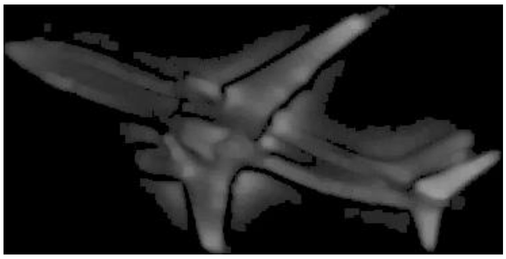

Fig. 7: A result of Seam Carving with Saliency Map energy allocation applied to Figure 5.

Fig. 8: SIFT-descriptors applied to Figure 7.

3.2 Flowchart of the Proposal

A flowchart of the proposal is shown in Figure 9. Our proposal is mainly based on the bag-of-keypoints (Bok) method, and newly applies Seam-carving and Saliency Map before the conventional Bok method.

1) Prepare learning images and test images.

2) Generate Saliency Map.

3) Carry out Seam-carving with Saliency Map energy allocation in the energy map calculation.

4) Generate feature vectors to all images with Bok method.

5) Learn or recognize an object by SVM (support vector machine).

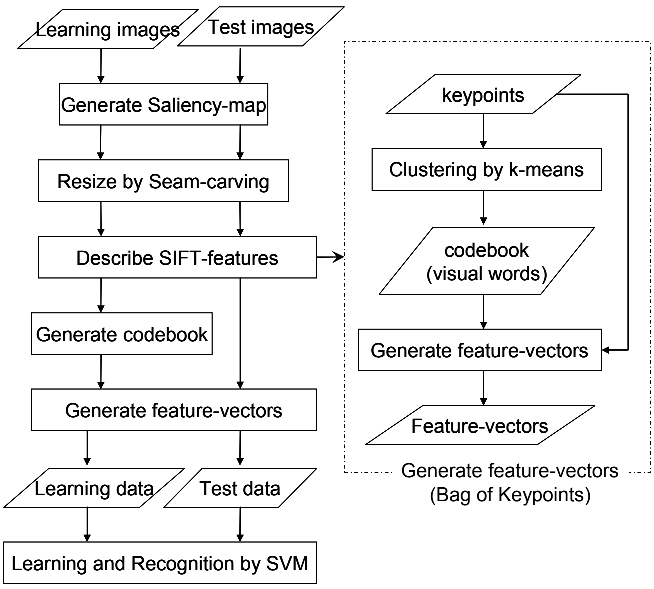

Fig.9 Flowchart of the proposal.

4. EXPERIMENTS

4.1 Experimentation environment

We used Caltech-101 that was a standard data set in this research field as an experimental data set. 10~101 kinds of categories were used from this data set. 20 images per category were used as a learning image, and 20 images per category were used as a test image.

Four types of evaluation tests were conducted.
①The first one is the conventional bag-of-features method.

②The second is a method applying Seam-carving and Saliency Map.

③The third is a method applying Seam-carving only.

④The fourth is a method applying Saliency Map, and limiting the description of keypoints only to the highly salient area.

4.2 Result

Figure 10 and Table 1 show an experiment result for Caltech-101 image dataset, in which a horizontal axis represents the number of classification categories and a vertical axis does performance gains of the three proposals against the Bok method.

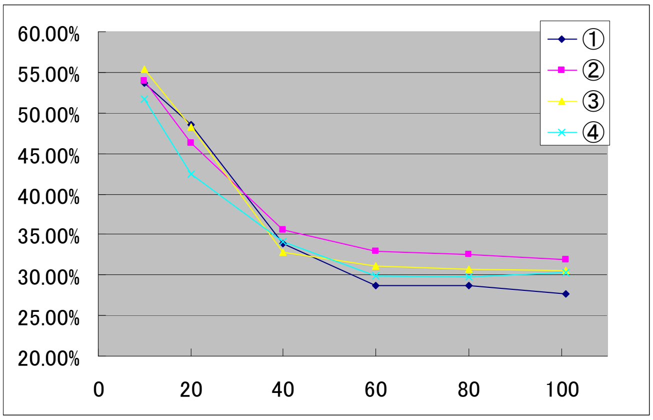

Fig.10 Experimental Result

Table 1 Experimental Result

Experimental method	①	②	③	④
10categories	53.65%	53.88%	55.42%	51.67%
20categories	48.56%	46.32%	48.19%	42.42%
40categories	33.90%	35.59%	32.83%	34.18%
60categories	28.71%	32.94%	31.15%	29.84%
80categories	28.67%	32.55%	30.63%	29.82%
101categories	27.72%	31.84%	30.61%	30.24%

4.2 Discussions

In comparison with the conventional technique, all proposal techniques showed good recognition rates. And, the method that combines Seam-carving and Saliency Map showed the highest identification rate than the other methods. An effect when the number of the categories increased became more remarkable. Any techniques did not have the difference at ten kinds of category classifications. But, at 101 kinds of category classifications, the proposal technique achieved a result that the identification rate was improved by about 3% compared to the conventional technique. Therefore, we can confirm the effect of our proposal technique.
The following is a reason why the identification rate was improved. At the dense sampling method, many SIFT descriptors will be described on non-object regions. When the category for recognition is few, the background area has important information for the object recognition.
However, we think that the importance of information of background regions decreases when the number of categories increases. And, we think that information on the object region can be learned more by limiting the SIFT descriptors of non-target regions, and the learning efficiency and the identification rate can be improved.

4. CONCLUSIONS

In this paper, we assume that the learning efficiency of images and the identification rate can be improved by removing areas that seem to be unnecessary for learning and identification as a preprocessing of generic object recognition. We used the Bok method for object recognition, SIFT for keypoint extraction, k-means clustering for feature vector quantization, and SVM for learning and recognition. Furthermore, in order to remove possible unnecessary regions, we applied Seam Carving and Saliency Map as pre-processing techniques before the Bok method. In the experiment, 10~101 kinds of categories were selected from Caltech-101 that was a standard data set in this research field and the multiclass classification was carried out. The result showed that our proposal can improve the identification rate by about 3% (10~20% at a part of categories) compared with the conventional technique.
As future tasks, we want to integrate a multiresolutional analysis technique used by both Saliency Map and SIFT into a single framework.

5. REFERENCES

[1] G. Csurka, C. Dance, L. Fan and C. Bray, "Visual categorization with bags of keypoints," ECCV, pp. 1-22, 2004.

[2] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, Vol.60, No.2, pp.91-110, Jan.2004.

[3] S. Avidan and A. Shamir, "Seam Carving for content-aware image resizing," SIGGRAPH, 2007.

[4] L. Itti, C. Koch and E. Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, No. 11, November 1998.