Contrast Context Histogram – A Discriminating Local Descriptor for Image Matching

Chun-Rong Huang[1,2], Chu-Song Chen[1] and Pau-Choo Chung[2]
1. Institute of Information Science, Academia Sinica, Taipei, Taiwan
2. Dept. of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan
{nckuos, song}@iis.sinica.edu.tw, pcchung@ee.ncku.edu.tw

Abstract

This paper presents a new invariant local descriptor, contrast context histogram, for image matching. It represents the contrast distributions of a local region, and serves as a local distinctive descriptor of this region. Object recognition can be considered as matching salient corners with similar contrast context histograms on two or more images in our work. Our experimental results show that the developed descriptor is accurate and efficient for matching.

1. Introduction

Image matching is difficult due to large variations of possible appearances of objects or scenes. The appearance variations might be caused by scale changes, rotations, different lighting conditions, and/or partial occlusions. Recently, invariant local descriptors constructed from images have been proposed to overcome appearance variations in object recognition [1][6][9]. The idea is to detect invariant local properties of salient image corners under a class of transformations, and then establish discriminating descriptors for these corners. In the past, Freeman and Adelson [3] developed steerable filters, which synthesize filters of arbitrary orientations from linear combinations of pixel derivatives in particular directions. Belongie *et al*. [1] proposed shape context that is a histogram of edge points with respect to a reference point under the log-polar coordinate. Ojala *et al*. [10] proposed a circularly symmetric binary pattern to discriminate textures. The gray value of the center point is subtracted by those of the local neighborhood points, and then quantized by a threshold to form a binary pattern. Lowe [6] proposed a scale invariant feature transformation (SIFT) descriptor that is invariant to scale and rotation. He computed discriminating image features through the detection of scale-space extremes. Then, invariant descriptors are constructed by using a weighted orientation histogram around the feature point.

In this paper, we propose a new invariant local descriptor called *contrast context histogram* (CCH) and apply it to image matching. CCH exploits contrast properties of a local region. Rotation and linear illumination changes are considered to make it robust against geometric and photometric transformations. In the experiments, we use CCH descriptors to represent clutter scenes and show its successfulness in image matching.

2. The CCH Descriptor

A main issue in developing invariant local descriptors is to represent a region effectively and discriminatively. Color histogram [2] is an option for textural description, but it is sensitive to illumination changes. Instead, we introduce the concept of contrast histogram to describe a component represented by an image patch, as shown below.

In our approach, we assume that there are already many salient corners extracted from an image I. For each salient corner p_c in the center of an $n \times n$ local region R, we compute the contrast $C(p)$ of a point p in R as

$$C(p) = I(p) - I(p_c), \qquad (1)$$

where $I(p)$ and $I(p_c)$ are the intensity values of p and p_c, respectively.

We then construct a descriptor of p_c based on these contrast values. In our approach, we separate R into several non-overlapping regions, R_1, R_2,…, R_t. Without lost of generality, we use a log-polar coordinate system (r, θ) to perform the division, as shown in Figure 1. Log-polar coordinate system has been used in many previous works [1][11], and is more sensitive to positions of nearby points respect to the center. To make the descriptor invariant to image rotations, the direction of $\theta = 0$ in the log-polar coordinate system is set to be coincide with the edge orientation of p_c.

How to represent a sub-region R_i efficiently and discriminatively is an important issue. We consider a histogram-based representation since histogram is

IEEE
COMPUTER
SOCIETY

relatively insensitive to non-uniform deformations of a region. An intuitive way to employ the histogram feature is to accumulate the contrast values in a sub-region into a histogram bin. However, summations of positive and negative contrast values may decrease the discriminating responses of the bin. Thus, to increase the discriminative ability of the descriptor, we introduce the positive and negative histogram bins of the contrast values for each sub-region, as shown in the following.

For each p in R_i, we define the positive contrast histogram bin with respective to p_c as

$$H_{R_i+}(p_c) = \frac{\sum \{ C(p) \mid p \in R_i \text{ and } C(p) \geq 0 \}}{\#_{R_i+}}, \quad (2)$$

where $\#_{R_i+}$ is the number of positive contrast values in R_i. In a similar manner, the negative contrast histogram bin is defined as

$$H_{R_i-}(p_c) = \frac{\sum \{ C(p) \mid p \in R_i \text{ and } C(p) < 0 \}}{\#_{R_i-}}, \quad (3)$$

where $\#_{R_i-}$ is the number of negative contrast values in R_i. By composing the contrast histograms of all the sub-regions into a single vector, the CCH descriptor of p_c with respect to its local region R is defined as follows:

$$CCH(p_c) = (H_{R_1+}, H_{R_1-}, ..., H_{R_t+}, H_{R_t-}). \quad (4)$$

To further overcome linear light changes, we normalize the CCH descriptor to a unit vector.

3. Experiments

We evaluated CCH descriptors on the data set[1] used in [9]. The dataset contains images with various geometric and photometric transformations for different scene types. To compute CCH descriptors from an input image, we first extracted corners from a multi-scale [5] Laplacian pyramid [8] by detecting Harris corners [4] on each level of the pyramid. A salient corner is selected if its minimal eigenvalue is larger than all the eigenvalues of its neighbors in a 7×7 region. Figure 1 illustrates the contrast context histogram of a salient corner p_c under the log-polar coordinate system. A local $n×n$ region R is divided into several sub-regions by quantizing r and θ of the log-polar coordinate system. For each sub-region, a 2-bin contrast histogram is constructed as introduced in
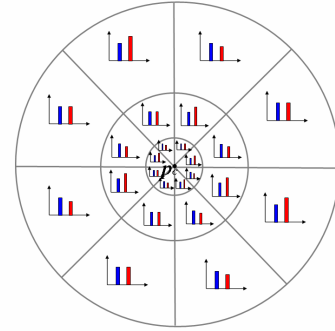
Figure 1. Log-polar diagram of the CCH descriptors. The center of the coordinate is the salient point p_c.

Section 2. A contrast context histogram descriptor of p_c is then computed as follows:

$$CCH(p_c) = (H_{r_0\theta_0+}, H_{r_0\theta_0-}, ..., H_{r_k\theta_{l-1}+}, H_{r_k\theta_{l-1}-}), \quad (5)$$

where $r_i = 0, ..., k, k = \lfloor \log(\sqrt{2n^2}) \rfloor, \theta_j = \frac{2\pi}{l}m, m = 0, ..., l-1$ and $CCH(p_c) \in R^{2(k+1)l}$.

Figure 2 shows several images of the data set, including rotation, viewpoint changes, image blur, JPEG compression, illumination changes, and zoom and rotation changes. For evaluation, we use recall as below:

$$recall = \frac{\#correct \ matches}{\#correspondences}. \quad (6)$$

For images of the same scene, one of them is chosen as a reference image, and the others are matched with it. Consider a pair of images (I, I'), where I is a reference image. For a corner p_c in I, let p_c' and p_c'' be the most similar and second-most similar corners to p_c in I', respectively, where the similarity is measured by the Euclidean distances of the CCH descriptors. Then (p_c, p_c') serves as a correspondence pair if

$$Dist(p_c, p_c') E_n < \alpha \times Dist(p_c, p_c''), \quad (7)$$

where $Dist(\cdot)$ is the Euclidean distance between the CCH descriptors and α is 0.6 in our experiment.

We compare CCH[2] with SIFT [6], which has been shown to be superior to other approaches in [9]. We used the SIFT code available in [7] in the experiments. Both CCH and SIFT were computed on images whose intensities are normalized to [0, 1]. We used $k = 3$ for the distance quantization and $l = 8$ for orientation quantization to implement the CCH descriptor according to empirical experience. Thus, the dimension of the CCH feature is 2×4×8 = 64. Figure 3

shows some examples of matching results by using CCH. The correspondences between two images were marked and connected with straight lines.

The results are shown in Table I. Both of these two methods have high matching accuracies on the dataset. Although the dimension of CCH is less than that of SIFT, it can perform comparable accuracy to SIFT.

Both methods were implemented by C# under an Intel P4 3.4G computer, and the computation time is shown in Table II. In this table, the descriptor time is the total time for salient corner selection and descriptor construction of the whole dataset. The matching time is the total time for finding the corresponding pairs of the whole dataset. It can be seen that the descriptor time of CCH is much less than that of SIFT because only subtraction is required when constructing CCH. By contrast, SIFT needs to compute magnitudes and the orientations of all pixels in a local sampled region. Average descriptor time is the average time to construct a descriptor for a salient corner. The matching time of CCH is also less because the dimension of CCH is smaller than that of SIFT. Our experiments show that the CCH descriptor has comparable matching accuracy than SIFT, but is more efficient to compute and match.

4. Summary

In this paper, we introduce CCH that is a new invariant descriptor to describe local properties of image patches. CCH is efficient to compute since simple subtraction is used in its construction. Employing positive and negative histogram bins of the contrast values makes CCH discriminative for image matching. Experimental results and comparative studies show that CCH has high matching accuracies with less computation time.

5. Acknowledgement

References

[1] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and Object Recognition Using Shape Contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no.4, pp. 509-522, 2002.

[2] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time Tracking of Non-rigid Objects Using Mean Shift," *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 142-151, 2000.

Table I. The matching results of CCH in the test data set.

Image set (Resolution)	CCH	SIFT
Rotation (512×512)	99.9%	99.7%
View change #1 (440×340)	100%	100%
View change #2 (400×320)	98.8%	98.2%
Image blur #1 (500×350)	99.7%	99.9%
Image blur #2 (500×350)	99.8%	99.9%
JPEG compression (400×320)	99.9%	99.9%
Illumination (460×306)	100%	99.7%
Rotation and zoom (420×336)	97.9%	98.9%
Over all	99.8%	99.8%
Total correspondences	7402	7228

Table II. The computation time of CCH and SIFT in the test data set.

	CCH	SIFT
(1) Descriptor time (s)	36.3	148.7
(2) Matching time (s)	66.5	128.6
Total running time (s) (1)+(2)	102.8	277.3
Average descriptor time (s)	0.0009	0.0035

[3] W. Freeman, and E. Adelson, "The Design and Use of Steerable Filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891-906, 1991.

[4] C. G. Harris, and M. J. Stephens, "A Combined Corner and Edge Detector", *In Proceedings of the Fourth Alvey Vision Conference,* pp.147-151, 1988.

[5] T. Lindeberg, "Scale-space Theory: A Basic Tool for Analyzing Structures at Different Scales," *Journal of Applied Statistic*, vol. 21, no. 2, pp. 224-270, 1994.

[6] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.

[7] LIBSIFT: http://user.cs.tu-berlin.de/~nowozin/libsift/

[8] K. Mikolajczyk, and C. Schmid, "Indexing Based on Scale Invariant Interest Points", *In Proceedings of International Conference on Computer Vision, ICCV'01*, vol. 1, pp. 525-531, 2001.

[9] K. Mikolajczyk, and C. Schmid, "A Performance Evaluation of Local Descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615-1630. 2005.

[10] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, 2002.

[11] S. Pereira, J.J.K.O. Ruanaidh, F. Deguillaume, G. Csurka, and T. Pun, "Template Based Recovery of Fourier-based Watermarks Using Log-polar and Log-log Maps, "*In Proceedings of Multimedia Computing and Systems*, vol. 1, pp. 870 – 874, 1999.
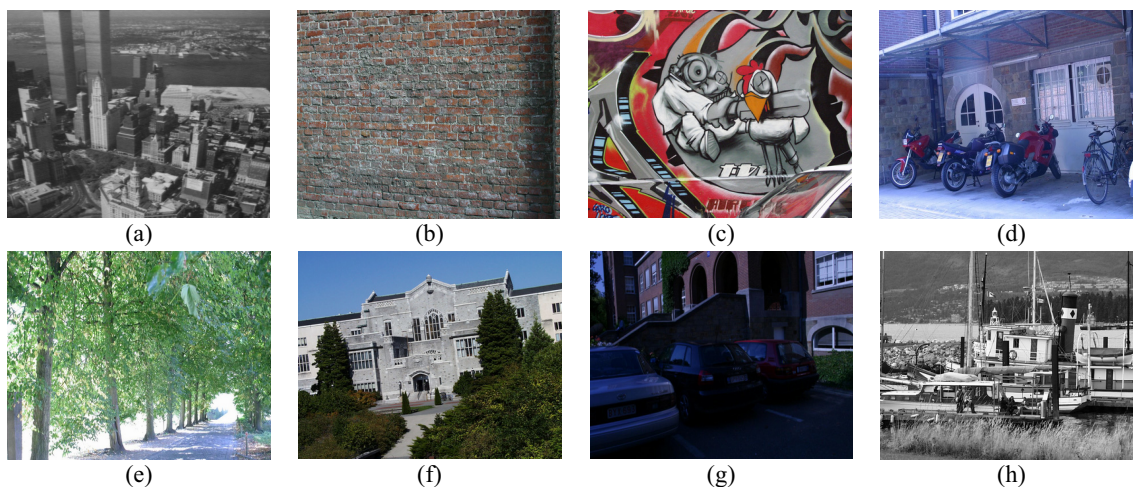
IEEE
COMPUTER
SOCIETY

Figure 2. Examples of images used for the evaluation under (a) Rotation, (b)(c) Viewpoint change, (d)(e) Image blur, (f) JPEG compression, (g) Light change and (h) Rotation and Zoom change.
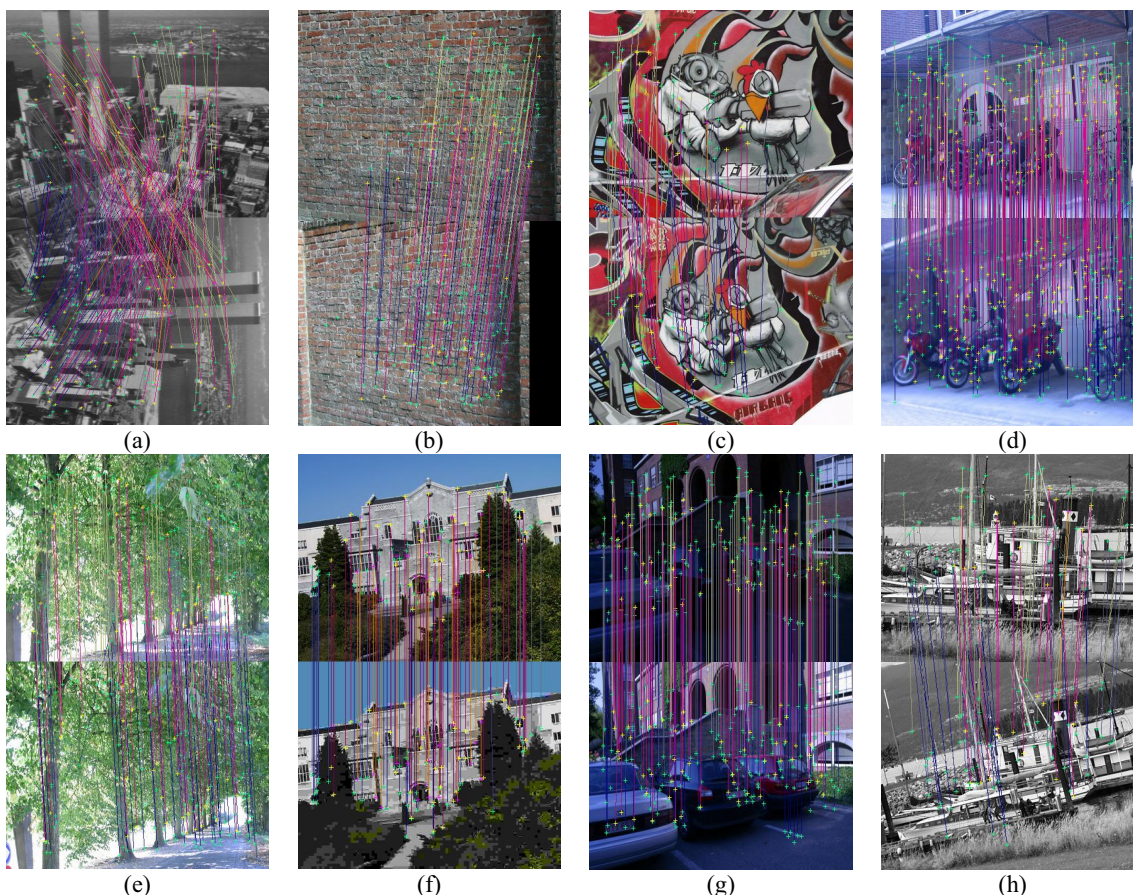

Figure 3. Examples of matching results under (a) Rotation, (b)(c) Viewpoint change, (d)(e) Image blur, (f) JPEG compression, (g) Light change and (h) Rotation and Zoom change.