Pattern Recognition 41 (2008) 3071 – 3077

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

# Contrast context histogram—An efficient discriminating local descriptor for object recognition and image matching

Chun-Rong Huang [a], Chu-Song Chen [a,*], Pau-Choo Chung [b]

[a] Institute of Information Science, Academia Sinica, Taipei 115, Taiwan
[b] Department of Electrical Engineering, National Cheng Kung University, Tainan 701, Taiwan

ARTICLE INFO

*Article history:*
Received 20 November 2006
Received in revised form 21 December 2007
Accepted 13 March 2008

*Keywords:*
Image matching
Local descriptor
Object recognition

ABSTRACT

In this paper, we propose a new invariant local descriptor, called the contrast context histogram (CCH), for image matching and object recognition. By representing the contrast distributions of a local region, it serves as a distinctive local descriptor of the region. Our experiments demonstrate that contrast-based local descriptors can represent local regions with more compact histogram bins. Because of its high matching accuracy and efficient computation, the CCH has the potential to be used in a number of real-time applications.



## 1. Introduction

Invariant local descriptors constructed from images have been proposed as a way to solve many problems, such as in image matching [1], object recognition [1–3], and many other vision-based applications [4–7]. The idea is to detect the invariant local properties of salient image corners under a class of transformations, and then establish discriminating descriptors for these corners. Descriptors provide robust representations of local image regions, even under partial occlusion. The basic problem is how to find the relevant information required to encode the local signatures.

Descriptors of local features have received considerable attention in recent years. For example, Freeman and Adelson [8] developed steerable filters, in which filters of arbitrary orientations are synthesized from linear combinations of pixel derivatives in particular directions. Belongie et al. [2] proposed a feature description called the shape context, which is a histogram of edge points with respect to a reference point under the log–polar coordinate. Lowe [1] introduced a scale invariant feature transformation (SIFT) descriptor that is invariant to scale and rotation. In this approach, keypoints are computed through the detection of scale-space extremes in a series of difference-of-Gaussian (DoG) images. Local descriptors are built up for each keypoint based on a weighted histogram of gradient orientations from a patch of pixels in its local neighborhood. In Refs. [3,9], it is shown that SIFT is one of the most effective matching approaches when scale and viewpoint changes occur in images. Various extensions of SIFT have been proposed. For example, Ke and Sukthankar proposed PCA-SIFT [10], which applies principal components analysis (PCA) [11] to a normalized gradient patch, instead of using SIFT's smoothed weighted histograms. The gradient location-orientation histogram (GLOH) [3] computes the SIFT descriptor for a log–polar location grid and then reduces the size of the descriptor with PCA. The primary focus of these extensions is to provide more distinctive and compact descriptors to improve the matching accuracy and speed.

In this paper, we propose a novel invariant local descriptor called the *contrast context histogram* (CCH) for image matching and object recognition. Our primary motivation is to develop a descriptor that is computationally fast, requires fewer histogram bins to represent a local region, and can achieve a good matching performance. CCH exploits the contrast properties of a local region, instead of storing the weighted edge orientation histograms of salient corners like the SIFT approach. Rotation and linear illumination changes are considered to make the CCH robust against geometric and photometric changes. Compared to the approaches such as SIFT (PCA-SIFT and GLOH) that require computing the gradient orientations of all the pixels in a region, CCH is more efficient to compute since it only evaluates the intensity differences between the center pixel and the other pixels in a region. Therefore, CCH is potentially more suitable for real-time applications such as augmented reality [5]. In the experiments, we use CCH descriptors to represent cluttered scenes and objects, and evaluate the method's effectiveness.

The remainder of the paper is organized as follows. In the next section, we describe the construction of CCH descriptors from the salient corners of images. Section 3 discusses the implementation of

* Corresponding author. Tel.: +886 227 883 799X1310.
*E-mail addresses:* nckuos@iis.sincia.edu.tw (C.-R. Huang), song@iis.sincia.edu.tw (C.-S. Chen), pcchung@ee.ncku.edu.tw (P.-C. Chung).

0031-3203/$30.00 © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2008.03.013

the CCH approach. The experiment results are reported in Section 4. Finally, in Section 5, we present our conclusions.

## 2. The CCH descriptor

The main issue in developing invariant local descriptors is how to represent a region effectively and discriminatively. The color histogram [12] is an option for textural description, but it is sensitive to illumination changes. Instead, we consider a technique that computes the contrast values of points within a region with respect to a salient corner. A contrast value is defined as the difference in intensity between a point and the salient corner. If the brightness of each pixel changes by adding a static value, the contrast will not be affected because it is computed from pixel differences.

Recently, Zabih and Woodfill proposed using rank and census transforms [13], which are insensitive to differences in camera gain or bias [14], to calculate the difference measures between all the pixels in a local region and a center pixel. In contrast, the local binary pattern (LBP) approach [15] only considers pixels on a circle of radius $R$ surrounding the center pixel. The gray value of the center point is subtracted from the value of the other pixels on the circle and quantized by a threshold to form a binary pattern. One limitation of these approaches is that the amount of information within a pixel is not very large. Moreover, only the sign of the intensity differences is used.

The proposed approach considers a histogram-based representation of the contrast values in the local region around the salient corners. In addition to considering the sign of the intensity differences, it uses the histogram values of the intensity differences to represent the appearance of objects around the salient points. In the following, we describe how to construct CCH descriptors.

We assume that several salient corners have been extracted from an image $I$ already. For each salient corner, $\boldsymbol{p}_c$, in the center of an $n \times n$ local region, $\boldsymbol{R}$, we compute the center-based contrast $C(\boldsymbol{p})$ of a point $\boldsymbol{p}$ in $\boldsymbol{R}$ as

$$C(\boldsymbol{p}) = I(\boldsymbol{p}) - I(\boldsymbol{p}_c), \tag{1}$$

where $I(\boldsymbol{p})$ and $I(\boldsymbol{p}_c)$ are the intensity values of $\boldsymbol{p}$ and $\boldsymbol{p}_c$, respectively.

We then construct a descriptor of $\boldsymbol{p}_c$ based on these contrast values. In our approach, $\boldsymbol{R}$ is separated into several non-overlapping regions, $\boldsymbol{R}_1, \boldsymbol{R}_2, \ldots, \boldsymbol{R}_t$. Without lost of generality, we use a log–polar coordinate system $(r, \theta)$ to perform the division, as shown in Fig. 1. The system, which has been used in many previous works [2,3,16], is more sensitive to the positions of points close to the center than to those of points farther away. To ensure that the descriptor is invariant to image rotations, the direction of $\theta = 0$ in the log–polar coordinate system is set to coincide with the edge orientation of $\boldsymbol{p}_c$.

Given the importance of representing a sub-region $\boldsymbol{R}_i$ efficiently and discriminatively, we consider a histogram-based representation, since histograms are relatively insensitive to non-uniform deformations of a region. An intuitive way to employ the histogram feature is to gather the contrast values in a sub-region into a histogram bin. However, the summations of the positive and negative contrast values may reduce the discriminative capability of the bin. Thus, to increase the discriminative ability of the descriptor, we introduce positive and negative histogram bins of the contrast values for each sub-region. We call them *contrast histograms*. Next, we define positive and negative contrast histograms.

For each $\boldsymbol{p}$ in $\boldsymbol{R}_i$, we define the positive contrast histogram bin with respect to $\boldsymbol{p}_c$ as

$$H_{\boldsymbol{R}_i+}(\boldsymbol{p}_c) = \frac{\sum\{C(\boldsymbol{p}) | \boldsymbol{p} \in \boldsymbol{R}_i \text{ and } C(\boldsymbol{p}) \geqslant 0\}}{\#_{\boldsymbol{R}_i+}}, \tag{2}$$

where $\#_{\boldsymbol{R}_i+}$ is the number of positive contrast values in the $i$-th region $\boldsymbol{R}_i$.

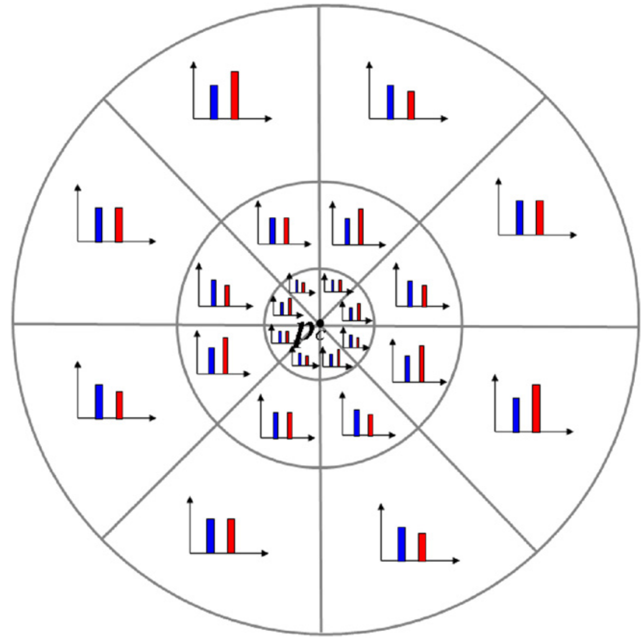



**Fig. 1.** Log–polar diagram of the CCH descriptors; the center of the coordinate is the salient point $\boldsymbol{p}_c$.

In a similar manner, the negative contrast histogram bin is defined as

$$H_{\boldsymbol{R}_i-}(\boldsymbol{p}_c) = \frac{\sum\{C(\boldsymbol{p}) | \boldsymbol{p} \in \boldsymbol{R}_i \text{ and } C(\boldsymbol{p}) < 0\}}{\#_{\boldsymbol{R}_i-}}, \tag{3}$$

where $\#_{\boldsymbol{R}_i-}$ is the number of negative contrast values in the $i$-th region $\boldsymbol{R}_i$.

By combining the values of all the contrast histogram entries from all the sub-regions into a single vector, the CCH descriptor of $p_c$ in association with its local region $\boldsymbol{R}$ can be defined as follows:

$$CCH(\boldsymbol{p}_c) = (H_{\boldsymbol{R}_1+}, H_{\boldsymbol{R}_1-}, H_{\boldsymbol{R}_2+}, H_{\boldsymbol{R}_2-}, \ldots, H_{\boldsymbol{R}_t+}, H_{\boldsymbol{R}_t-}), \tag{4}$$

which can also be considered as signed measurements of local contrasts with respect to a salient point $\boldsymbol{p}_c$. The vector length of the descriptor corresponds to the number of histogram bins, as shown in Fig. 1. The figure illustrates an example of two histogram bins in $3 \times 8 = 24$ sub-regions; hence, there are 48 elements in the descriptor. To deal with linear lighting changes, we normalize the CCH descriptor to a unit vector, which resolves the problem of changes in contrast due to multiplying a constant.

## 3. Implementation

To compute CCH descriptors from an input image, we first extract the corners from a multi-scale Laplacian pyramid [17] by detecting the Harris corners [18] on each level of the pyramid. As noted in Ref. [17], corners that are invariant to the scale changes of the image can be detected by searching for stable features on Laplacian pyramids in the scale space [19]. A salient corner is selected if its minimal eigenvalue is larger than all the eigenvalues of its neighbors in a $7 \times 7$ region. Fig. 1 illustrates the CCH of a salient corner $\boldsymbol{p}_c = (x_c, y_c)$ under the log–polar coordinate system.

A local region $\boldsymbol{R}$ surrounding $\boldsymbol{p}_c$ is divided into several sub-regions by quantizing $r$ and $\theta$ of the log–polar coordinate system as follows:

$$r = \log(\sqrt{(x - x_c)^2 + (y - y_c)^2}) \tag{5}$$

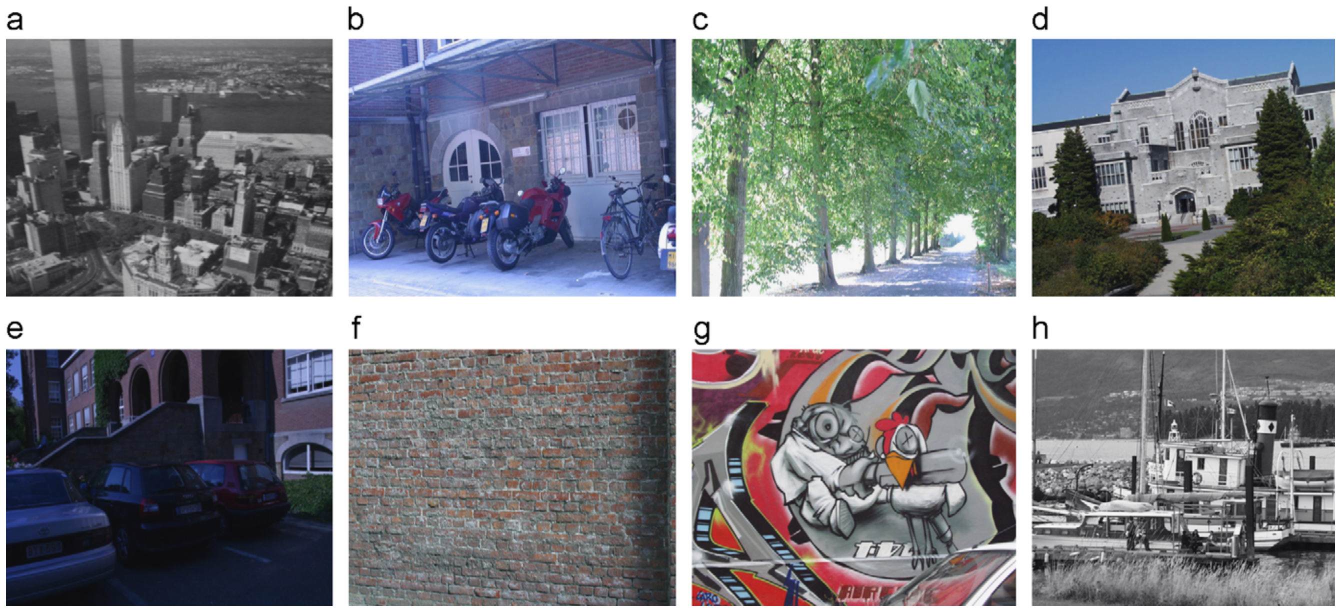


**Fig. 2.** Images used for evaluation under (a) rotation, (b) and (c) image blurring, (d) JPEG compression, (e) lighting changes, (f) and (g) viewpoint changes, and (h) rotation and zoom changes.

and

$$\theta = \tan^{-1}\left(\frac{y - y_c}{x - x_c}\right). \tag{6}$$

For each sub-region, a 2-bin contrast histogram is constructed using the method described in Section 2. A CCH descriptor of $\mathbf{p}_c$ is then computed as follows:

$$CCH(\mathbf{p}_c) = (H_{r_0\theta_0+}, H_{r_0\theta_0-}, \ldots, H_{r_k\theta_{l-1}+}, H_{r_k\theta_{l-1}-}), \tag{7}$$

where $r_i = 0, \ldots, k$, $k = \lfloor \log(\sqrt{2n^2}) \rfloor$, $\theta_j = (2\pi/l)m$, $m = 0, \ldots, l-1$ and $CCH(\mathbf{p}_c) \in R^{2(k+1)l}$. To implement the CCH descriptor,[1] we suggest $k = 3$ for the distance quantization and $l = 8$ for the orientation quantization. Thus, the dimensions of the CCH descriptor are $2 \times 4 \times 8 = 64$.

## 4. Experiments

### 4.1. Data set

We evaluated CCH descriptors on the data set[2] used in Ref. [3]. It contains images of various geometric and photometric transformations for different scene types. Fig. 2 shows the following images extracted from the data set: rotation, image blur, JPEG compression, lighting changes, viewpoint changes, and zoom and rotation changes. In the case of rotation, the images were obtained by rotating the camera around its optical axis, and the blurred sequences were acquired by varying the camera's focus. The JPEG sequence was generated by varying the image quality parameter of JPEG, and the lighting change images were derived by varying the camera's aperture. To capture the viewpoint changes, the camera's position was varied from a front view to a side view at approximately 60°. The zoom and rotation changes were obtained by varying the camera's focus and rotation simultaneously. Since the images are either of planar scenes or the camera position was fixed during acquisition, we can consider that they are related by homographies (plane projective transformations). In the data set, ground truth homographies are obtained by robust a small-baseline homography estimation algorithm that automatically detects and matches points of interest [20].

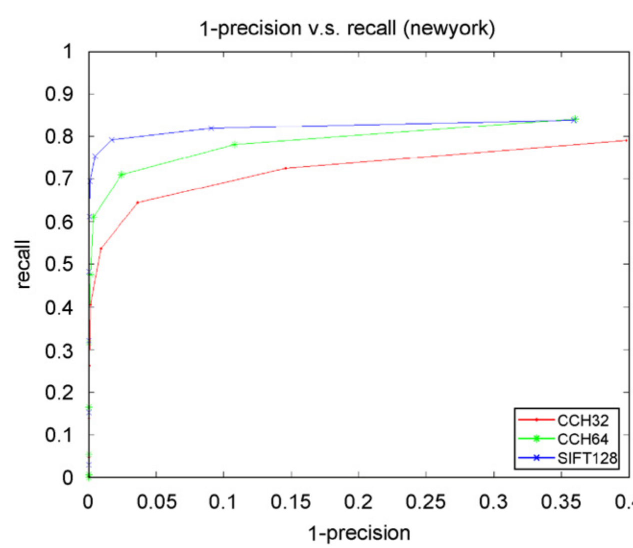


**Fig. 3.** Evaluation of in-plane image rotation.

### 4.2. Experiment setup

To evaluate the image matching performance, we use the criterion proposed in Ref. [3], which is based on the number of correct matches and false matches obtained for a pair of images. Consider an image pair $(I, I')$, where $I$ is the reference image and $I'$ is the test image. For a corner $\mathbf{p}_c$ in $I$, let $\mathbf{p}'_c$ and $\mathbf{p}''_c$ be, respectively, the most similar and second-most similar corners to $\mathbf{p}_c$ in $I'$. The similarity between the corners is measured by the Euclidean distance between

[1] The test software for CCH is available at http://imp.iis.sinica.edu.tw/CCH/CCH.htm

[2] The data set is available at http://www.robots.ox.ac.uk/~vgg/research/affine

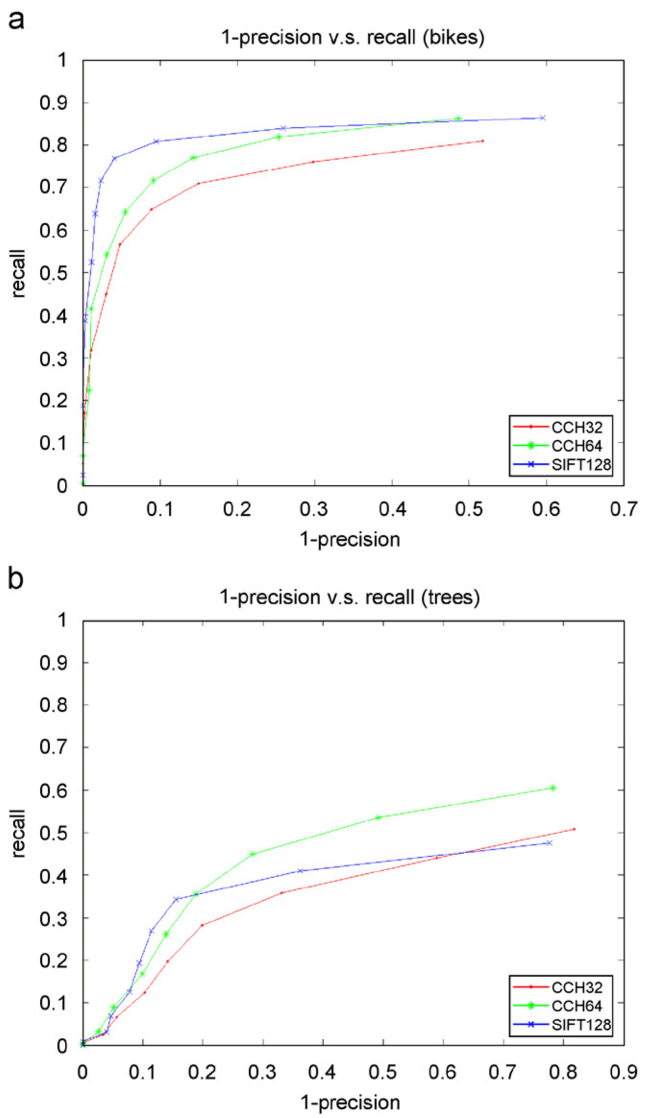


**Fig. 4.** Evaluation of blurred images of (a) a structured scene and (b) a textured scene.

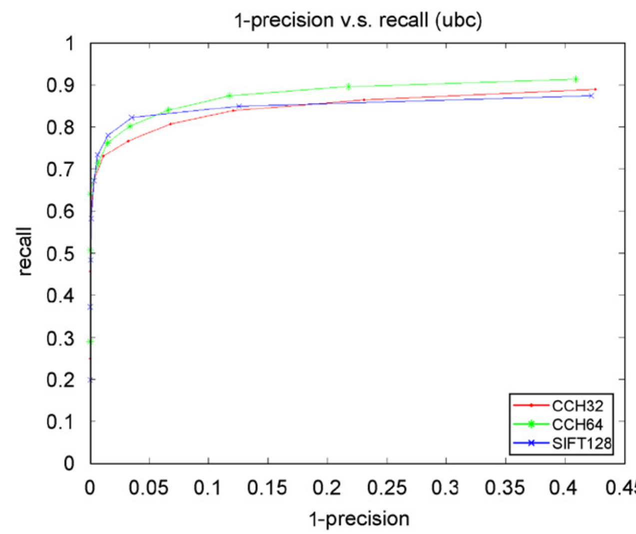


**Fig. 5.** Evaluation of JPEG compression.

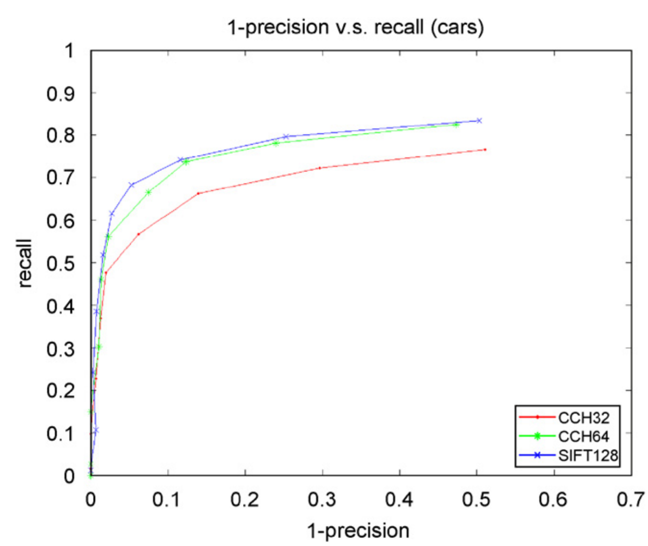


**Fig. 6.** Evaluation of illumination changes.

the descriptors of the corners. As with SIFT [1], $\boldsymbol{p}_c$ and $\boldsymbol{p}'_c$ match if

$$Dist(\boldsymbol{p}_c, \boldsymbol{p}'_c) < \alpha \times Dist(\boldsymbol{p}_c, p''_c), \tag{8}$$

where $Dist(\cdot)$ is the Euclidean distance between the descriptors. Each descriptor in the reference image is compared with each descriptor in the test image, after which we count the number of correct matches and the number of false matches. The value of $\alpha$ is varied to obtain the curves, and the results are presented with recall versus 1-precision. Recall is the ratio of the number of correctly matched corners to that of the corresponding corners between two images of the same scene:

$$\text{recall} = \frac{\#\text{ of correct matches}}{\#\text{ of correspondences}}. \tag{9}$$

To verify the correct matches, we use the relative locations and overlap errors of the salient corners [3]. The relative location measures how well the salient corners correspond under a particular transformation, which is a homography in our case. The relative location $l$ is defined as $\|\boldsymbol{p}'_c - \boldsymbol{H}\boldsymbol{p}_c\|$, where $\boldsymbol{H}$ is the homography between the reference and the test image. Moreover, we consider the overlap errors in the area of an image covered by two corresponding neighborhoods points. The overlap error, $\varepsilon$, is calculated by the ratio of the intersection over the union of the regions as follows [3]:

$$\varepsilon = 1 - \frac{\boldsymbol{A} \cap \boldsymbol{H}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{H}}{\boldsymbol{A} \cup \boldsymbol{H}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{H}}, \tag{10}$$

where A and B are the regions near $\boldsymbol{p}_c$ and $\boldsymbol{p}'_c$, respectively. A match is correct if the distance between the estimated and real locations, $l$, is less than 4, and the overlap error in the image area covered by two corresponding regions is less than 50 percent of the overall region. The number of correspondences (possible correct matches) is determined by the same criterion.

The number of false matches relative to the total number of matches is represented by 1-precision as follows:

$$\text{1-precision} = \frac{\#\text{ of false matches}}{\#\text{ of correct matches} + \#\text{ of false matches}}. \tag{11}$$

Note that recall and 1-precision are independent terms. Recall is computed with respect to the number of corresponding salient corners, while 1-precision is computed with respect to the total number of matches using the matching criteria in Eq. (8). As noted in

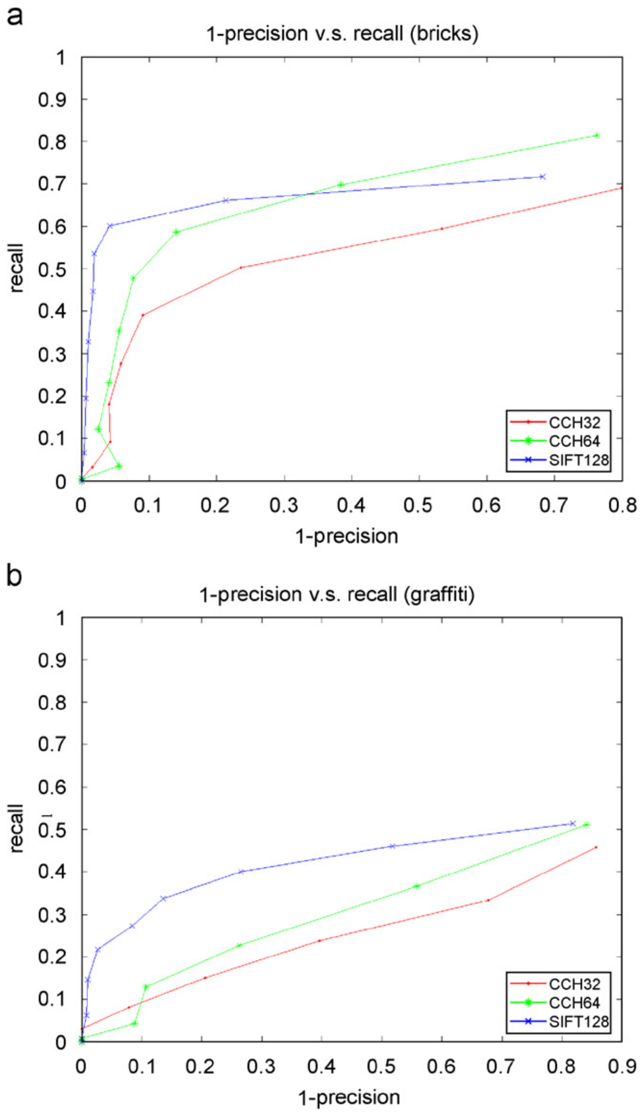


**Fig. 7.** Evaluation of viewpoint changes for (a) a structured scene and (b) a textured scene.

Ref. [3], a perfect descriptor would yield a recall rate equal to 1 for any precision rate; however, in practice, the recall increases as the value of $\alpha$ increases. In our experiments, we computed the average recall and 1-precision rates for each transformation from all the test images with respect to the reference image to construct the recall versus the 1-precistion curves.

The performance of CCH is compared with SIFT [1], which has been shown to be superior to the approaches in Refs. [3,9]. We used the Windows execute file[3] of SIFT in the experiments. Both methods were implemented in the same language, C, and the programs were run on an Intel P4 3.4 G computer. Note that CCH descriptors are computed on images whose intensities are normalized in the interval range [0, 1].

We suggest that the dimension of the CCH descriptor should be 64 under the log–polar coordinate system, denoted as CCH64 in the following experiments. For comparison, like SIFT, we also applied the CCH descriptor to the $4 \times 4$ sample regions surrounding salient corners. This descriptor, denoted as CCH32, is comprised of $2 \times 4 \times 4 = 32$ elements. In our experiments, we applied the SIFT descriptor with dimension 128 as suggested in Ref. [1], which is commonly used in many applications.

[3] The software of SIFT is available at http://www.cs.ubc.ca/~lowe/keypoints/

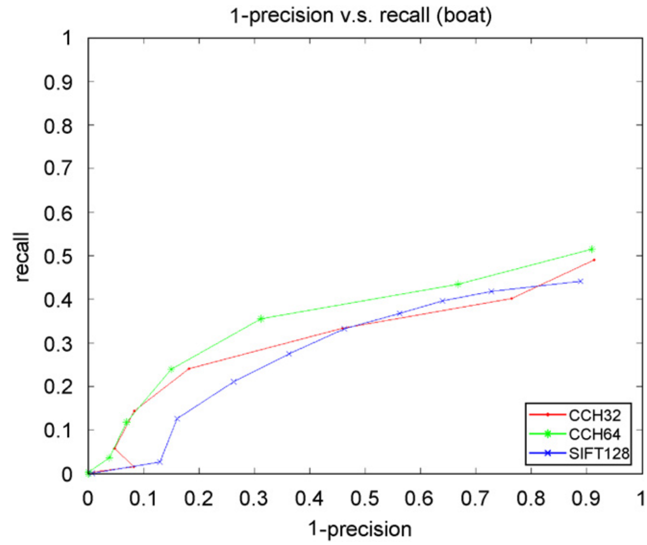


**Fig. 8.** Evaluation of scale and rotation changes.

**Table 1**
The computation times of CCH and SIFT for the test data set (48 images)

| Time (s) | CCH32 | CCH64 | SIFT |
|---|---|---|---|
| (1) Descriptor | 9 | 16 | 84 |
| (2) Matching | 18 | 35 | 69 |

### 4.3. Experiment results

In this section, we present the results of our evaluation. The first experiment tests the robustness of the local descriptor against in-plane rotation by rotating the images between 10° and 40°. In Fig. 3, the $x$-axis represents the 1-precision and $y$-axis represents the recall rates of different descriptors, and SIFT128 denotes the SIFT descriptor of dimensional 128. All curves are nearly horizontal at similar recall values, i.e., in-plane rotation does not affect these descriptors. CCH64 and SIFT are better than CCH32. This result implies that the higher dimensions of the descriptor are better at distinguishing between images under in-plane rotation.

We now test the descriptors on natural scenes that contain a significant amount of blur (change of camera focus). Figs. 4(a) and (b) show the results for a structured scene and a textured scene, respectively. Although varying the camera focus changes the pixel intensities and the edges of the images, the descriptors achieve high recall rates on the structured scene; however, the recall rates on the textured scene are lower. The performance of CCH64 is comparable with that of SIFT for the structured scene, and better than SIFT for the textured scene.

JPEG compression is widely used to modify images in real applications. Matching descriptors with JPEG images indicates whether the descriptors can still represent image regions under compression. As shown in Fig. 5, the recall rates are evenly high for all descriptors, especially CCH64.

Fig. 6 shows the results of illumination changes caused by varying the camera aperture. Because CCH and SIFT consider linear illumination changes, the effect of such changes on an image is mitigated. As a result, the recall rates are satisfactory. Moreover, CCH64 and SIFT exhibit similar robustness against changes in lighting conditions.

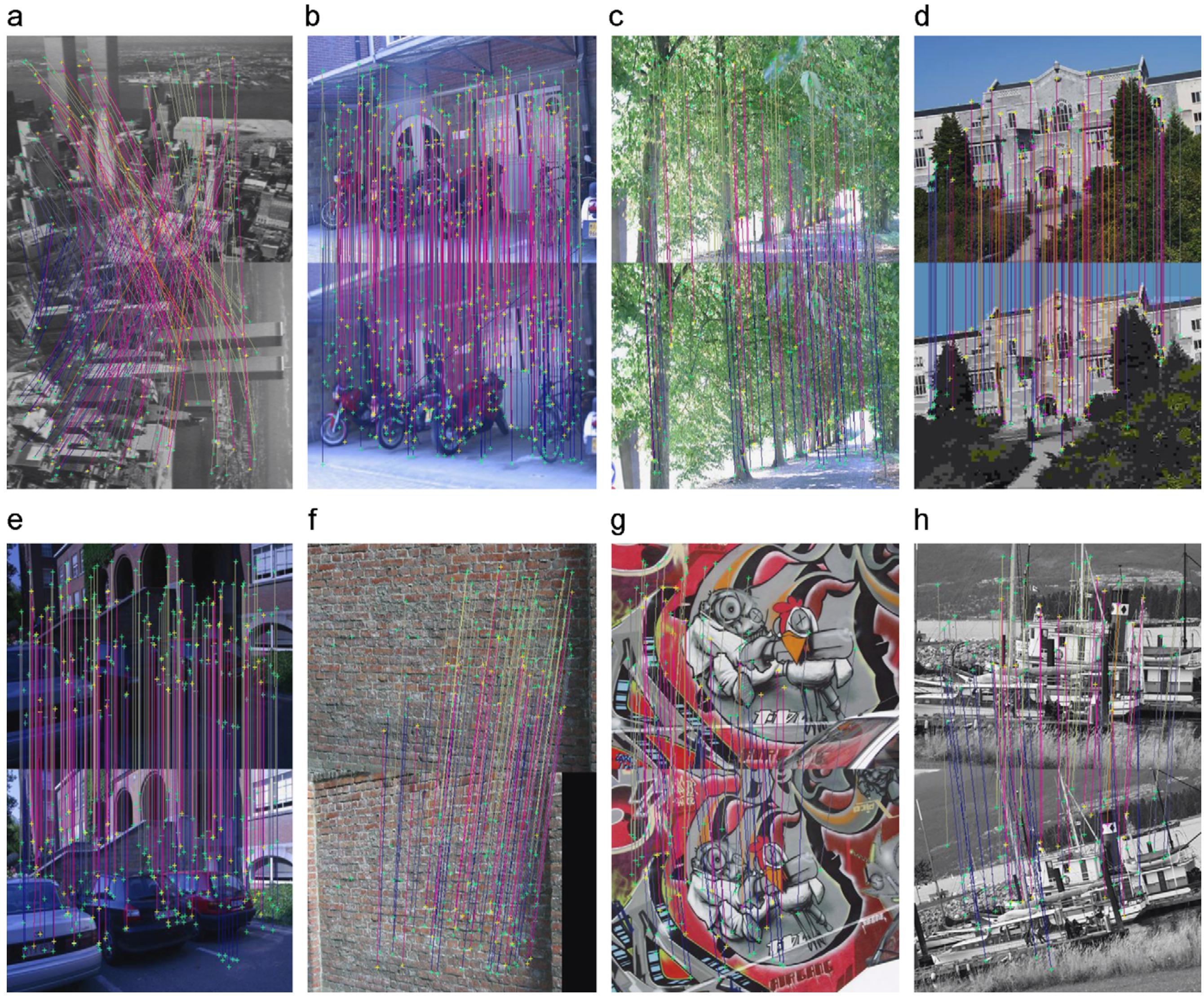


**Fig. 9.** Examples of matching results under (a) rotation, (b) and (c) image blurring, (d) JPEG compression, (e) lighting changes, (f) and (g) viewpoint changes, and (h) rotation and zoom changes.

In contrast, the performance of CCH32 is poor due to the limited length of the descriptor.

Next, we evaluate the performance based on viewpoint changes from the front view angle to approximately 50°. As noted in Ref. [3], there are also some scale and brightness changes in the test images. Figs. 7(a) and (b) show the results for the structured scene in Fig. 2(f) and the textured scene in Fig. 2(g), respectively. Similar to the case of blurred images, the recall rates of all descriptors are better for the structured scene. We found that the salient corners can be detected easily in the structured scene. The images are also easier to match because the blocks are obviously from different near-regular regions, which can capture sufficiently distinctive signal variations. The textured scene contains rich edge information and large regions with similar colors; thus, the edge-based descriptor achieves better results.

Scale changes with camera rotations cause substantial changes in an object's appearance due to the perspective projection. Scale changes lie in the range 1–2.5, while image rotations are in the range 30–45°. From Fig. 8, we observe that the performance of all descriptors is affected by scale changes with camera rotations, even though we impose the scale space to detect salient corners. CCH achieved better results than SIFT, possibly because the repeatability of the Harris–Laplacian corner is superior to the DOG corner [17]. The Harris–Laplacian corner can be detected better under scale and rotation changes. Therefore, we selected the Harris–Laplacian corner in our approach.

Table 1 lists the computation times for CCH and SIFT for the whole data set (48 images). The descriptor time is the total time taken for salient corner selection and descriptor construction of the whole data set. The matching time is the total time required to find the corresponding pairs in the whole data set. The descriptor time of CCH is much less than that of SIFT because only subtraction is required to construct CCH. In contrast, SIFT needs to compute the magnitudes and orientations of all the pixels in a locally sampled region. The matching time of CCH is also lower because the dimensions of CCH are smaller than those of SIFT. Our experiments show that, although the accuracy of the CCH and SIFT descriptors is comparable, CCH is more efficient in constructing descriptors and matching images.

The correspondences between two images are marked and connected with straight lines. For images of the same scene, one is chosen as a reference image and the others are used as test images. Fig. 9 shows some examples of the matching results using CCH.

## 5. Conclusion

We have proposed a new invariant descriptor called CCH to describe the local properties of image patches, and shown that it is computationally efficient and highly effective in determining the correspondences between images. It is successful because the positive and negative histogram bins of the contrast values are discriminative properties of local regions that can be computed rapidly as their construction only involves simple subtractions. The experiment results suggest that CCH has considerable potential to be used for real-time image-matching and object-recognition applications. In our future work, we will extend the gray-value-based CCH to a chromatic-based version so that more discriminative descriptors can be applied to color images.

## Acknowledgments

This research was supported in part by NSC 96-3113-H-001-011 and NSC 96-2752-E-002-007-PAE from the National Science Council, Taiwan. They would also like to thank Mr. P. Dunne for his work on polishing the writing.

## References

[1] D. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vision 60 (2) (2004) 91–110.
[2] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, IEEE Trans. Pattern Anal. Mach. Intell. 24 (4) (2002) 509–522.
[3] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, IEEE Trans. Pattern Anal. Mach. Intell. 27 (10) (2005) 1615–1630.
[4] T. Ahonen, A. Hadid, M. Pietikäinen, Face recognition with local binary patterns, in: Proceedings of the European Conference on Computer Vision, ECCV'04, vol. 1, 2004, pp. 469–481.
[5] I. Skrypnyk, D.G. Lowe, Scene modeling, recognition and tracking with invariant image features, in: Proceedings of International Symposium on Mixed and Augmented Reality, ISMAR'04, 2004, pp. 110–119.
[6] J. Sivic, A. Zisserman, Video google: a text retrieval approach to object matching in videos, in: Proceedings of the International Conference on Computer Vision, ICCV'03, vol. 2, 2003, pp. 1470–1477.
[7] A. Thayananthan, B. Stenger, P. Torr, R. Cipolla, Shape context and chamfer matching in cluttered scenes, in: Proceedings of Computer Vision and Pattern Recognition, CVPR'03, vol. 1, 2003, pp. 127–133.
[8] W. Freeman, E. Adelson, The design and use of steerable filters, IEEE Trans. Pattern Anal. Mach. Intell. 13 (9) (1991) 891–906.
[9] P. Moreels, P. Perona, Evaluation of features detectors and descriptors based on 3D objects, in: Proceedings of International Conference on Computer Vision, ICCV'05, vol. 1, 2005, pp. 800–807.
[10] Y. Ke, R. Sukthankar, PCA-SIFT: a more distinctive representation for local image descriptors, in: Proceedings of Computer Vision and Pattern Recognition, CVPR'04, vol. 1, 2004, pp. 506–513.
[11] R. Duda, P. Hart, D. Stork, Pattern Classification, Wiley, New York, 2001.
[12] D. Comaniciu, V. Ramesh, P. Meer, Real-time tracking of non-rigid objects using mean shift, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition CVPR'00, vol. 2, 2000, pp. 142–149.
[13] R. Zabih, J. Woodfill, Non-parametric local transforms for computing visual correspondence, in: Proceedings of the Third European Conference on Computer Vision, ECCV'94, vol. 2, 1994, pp. 151–158.
[14] D. Scharstein, R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, Int. J. Comput. Vision 47 (1/2/3) (2002) 7–42.
[15] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 971–987.
[16] S. Pereira, J.J.K.O. Ruanaidh, F. Deguillaume, G. Csurka, 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, 1999, pp. 870–874.
[17] K. Mikolajczyk, C. Schmid, Indexing based on scale invariant interest points, in: Proceedings of International Conference on Computer Vision, ICCV'01, vol. 1, 2001, pp. 525–531.
[18] C.G. Harris, M.J. Stephens, A combined corner and edge detector, in: Proceedings of the Fourth Alvey Vision Conference, 1988, pp. 147–151.
[19] T. Lindeberg, Scale-space theory: a basic tool for analyzing structures at different scales, J. Appl. Stat. 21 (2) (1994) 224–270.
[20] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, Cambridge, 2000.

**About the Author**—CHUN-RONG HUANG received the B.S. degree in electrical engineering from National Cheng Kung University, Taiwan, in 1999. He received the Ph.D. in Electrical Engineering from National Cheng Kung University, Taiwan, in 2005. He is currently a postdoctoral fellow in the Institute of Information Science, Academia Sinica, Taipei, Taiwan. His research interests include computer vision, computer graphics, and medical image processing.

**About the Author**—CHU-SONG CHEN received a B.S. in control engineering from National Chiao-Tung University, Hsing-Chu, Taiwan, in 1989. He received an M.S. in 1991 and a Ph.D. in 1996, respectively, both from the Department of Computer Science and Information Engineering, National Taiwan University. He is now an associate research fellow of the Institute of Information Science, Academia Sinica, and also an adjunct associate professor of Graduate Institute of Networking and Multimedia, National Taiwan University. His research interests include computer vision, pattern recognition, signal/image processing, and multimedia. Since 2007, Dr. Chen serves as the Secretary-General of the Image Processing and Pattern Recognition (IPPR) Society, Taiwan, which is one of the societies of the International Association of Pattern Recognition (IAPR). He has published more than 70 technical papers, and has received the outstanding paper awards of IPPR in 1997, 2001, and 2005.

**About the Author**—PAU-CHOO CHUNG received the B.S. and M.S. degrees in electrical engineering from National Cheng Kung University, Taiwan, Republic of China, in 1981 and 1983, respectively, and the Ph.D. degree in electrical engineering from Texas Tech University, in 1991. In 1991, she joined the Department of Electrical Engineering, National Cheng Kung University, and has become a full professor since 1996. Since 2001, she has served as the vice director, and currently the Director, of the Center for Research of E-life Digital Technology, National Cheng Kung University. She was selected as Distinguished Professor of National Cheng Kung University in 2005. She also served as the Director of Electrical Laboratory, National Cheng Kung University, during 2005–2008.
Dr. Chung's research interests include image analysis and pattern recognition, video image analysis, neural networks, telemedicine, and multimedia processing. Particularly she applies most of her research results on medical applications and received many awards. Dr. Chung has served as the program committee member in many international conferences. She is currently the chair of the IEEE Life Science Systems and Applications Technical Committee. She is also a member of the IEEE Visual Signal Processing and Communication Technical Committee, the IEEE Neural Systems and Applications Technical Committee, and the Multimedia Systems & Applications Technical Committee in the CASS. Currently she is also serving as the Associate Editor of Journal of Information Science and Engineering, the Guest Editor of IEEE Transactions on Circuits and Systems-I, a Member on IEEE International Steering Committee, IEEE Asian Pacific Conference on Circuits and Systems. She was the Chair of IEEE Computational Intelligence Society, Tainan Chapter (2005–2006), the Secretary General of Biomedical Engineering Society of the Republic of China (2005–2006), and Delegate in International Federation for Medical and Biological Engineering (IFMBE). She is one of the co-founders of Medical Image Standard Association at Taiwan. She is currently a member in BoG of CAS Society (2007–2009).
She is a member of Phi Tau Phi honor society and IEEE Fellow.