Keypoint Recognition with Histograms of Normalized Colors

Clark F. Olson and Siqi Zhang
Computing and Software Systems, School of STEM
University of Washington Bothell, USA
Email: cfolson@uw.edu

***Abstract*—Keypoint recognition is an important component of many object recognition and image classification systems. However, the color information present in images is not well incorporated. Previous techniques mostly rely on derivatives that only use the rate of change, rather than the actual color, and/or neglect the spatial relationships between the colors. We describe a new keypoint descriptor constructed with normalized color histograms that preserves color information and spatial relationships while maintaining invariance to illumination brightness variation (bias and gain). When combined with shape information, our descriptor surpasses the performance of previous techniques.**

***Keywords*-object recognition; keypoint; descriptor; color;**

I. Introduction

Keypoint detection and recognition has a long history in computer vision for image matching, object recognition, and image classification. Considerable work on keypoint descriptors has focused on the use of invariants based on gradients in grayscale and color images [1], [2], [3], [4]. Less work has considered the underlying color values (rather than the derivatives). The use of local hue, for example, has been examined [5], [6], [4].

In this paper, we consider a descriptor that accumulates color values directly (after appropriate normalization) in order to retain as much information as possible. Our technique, Histograms of Normalized Colors (HoNC), uses simple color histograms (not gradients) in a spatial grid, similar to the manner in which SIFT uses gradient orientation histograms to describe a keypoint. With normalization, our descriptor is invariant to illumination intensity changes (bias and gain), but not illumination color. With a modified normalization, it can be made invariant to illumination color as well. However, this removes information that is useful for keypoint recognition when the illumination color is unchanged.

We note that, in a previous study on color descriptors [4], few descriptors were invariant to illumination color (including SIFT). Among those that were, only RGB-SIFT (SIFT descriptors computed independently on the three color channels and then combined) was competitive with the top performing descriptors. Furthermore, Zhang et al. [7] argue that "local features with the highest possible level of invariance do not yield the best performance." The reason for this is that a feature loses discriminative power when it is invariant to phenomena that it does not need to be.

In addition, techniques based on pixel hues or ratios of colors that remove intensity information cannot succeed on purely grayscale images, since all pixels yield identical hue values (or color ratios). Indeed, any image with low saturation will cause these techniques problems. Our technique is hindered by grayscale images, but it still yields a useful descriptor.

Experiments on the Oxford affine covariant regions data set used in previous studies [8], [9] and additional image pairs indicate that the HoNC descriptor performs comparably to the SIFT descriptor and its variations. When combined with SIFT, performance gains are achieved.

The next section discusses previous work. Section III describes our new descriptor (HoNC). Section IV gives details on competing (and complementary) descriptors. Section V describes our experiments comparing HoNC with other techniques. Finally, Section VI gives our conclusions.

II. Previous Work

Early work on keypoint detection focused on corners [10], [11]. Schmid and Mohr [1] first developed invariant descriptors for keypoints that allowed them to be recognized under arbitrary rotation and scale. Their work was based on the local jet at a point in the image and was computed using the average intensity and derivatives up to the third order.

Lowe's work [2] has been the most influential on keypoint recognition. Keypoints were first detected by finding scale-space extrema after applying difference-of-Gaussian (DoG) operations to an image. This yielded the keypoint location and size. Orientation of the keypoint was determined using a histogram of local gradient orientations.

Next, the Scale-Invariant Feature Transform (SIFT) descriptor was created by (implicitly) placing a 4×4 grid at the keypoint location, scale, and orientation. In each grid cell, an eight-bin histogram of gradient orientations (weighted by the gradient magnitude) was constructed, resulting in a $4 \times 4 \times 8 = 128$ dimensional descriptor for each keypoint. This keypoint descriptor is invariant to keypoint location, scale, orientation, and affine illumination changes. Keypoint descriptors can be easily compared using the Euclidean distance to determine similarity. An example of SIFT matching

978-1-5090-2491-9/16 $31.00 © 2016 IEEE
DOI 10.1109/CRV.2016.37

Figure 1. Example of SIFT keypoint matching on images of furry creatures. The top 200 matches are displayed. With SIFT matching six incorrect matches are found.

(using the OpenCV implementation based on the work of Rob Hess [12]) can be seen in Fig. 1.

Many variations and improvements on SIFT for grayscale image data were subsequently developed, including HOG [13], GLOH [9], SURF [14], BRIEF [15], and ORB [16].

Less work has considered the use of color data. Among the first to use color information in keypoint descriptors were van de Weijer and Schmid [5]. They developed multiple descriptors, including a hue histogram (invariant to illumination intensity, but not illumination color) and opponent angle histogram, which has additional invariance to diffuse lighting. However, the color histograms have no spatial component, based on the argument that combining the descriptor with SIFT provides sufficient spatial information.

Abdel-Hakim and Farag [17] describe the CSIFT descriptor based on color invariants. They use a color invariant from Guesebroek et al. [18] combined with the SIFT descriptor (replacing intensity with the color invariant) for keypoint description. Burghouts and Geusebroek [3] use similar descriptors based on color invariants. Subsequently, van de Sande et al. [4] noted that the CSIFT descriptor can be seen as SIFT using opponent color space coordinates normalized by intensity.

Luke, Keller, and Chamorro-Martinez [6] describe a descriptor using hue combined with spatial information. Their descriptor replaces gradient orientation and magnitude in the SIFT descriptor with pixel hue and saturation. The descriptor is then stacked with the original SIFT descriptor.

Several color descriptors were compared by van de Sande et al. [4]. The four best performing descriptors in their experiments are RGB-SIFT (SIFT on all three color channels separately, then stacked), OpponentSIFT (same as RGB-SIFT, but in an opponent color space), C-SIFT (same as OpponentSIFT, but with the first two components divided by the third component, i.e., intensity), and rgSIFT (same as RGB-SIFT, but with only red and green divided intensity). In their experiments, descriptors based on simple color histograms fared poorly. However, they contained no spatial information.

III. Histograms of Normalized Colors

Our contribution is a new descriptor that combines color and spatial information. It is invariant to affine illumination brightness changes (bias and gain) without resorting to gradient information and it can be applied to images with little (or no) saturation.

The premise of our descriptor is to compute normalized color histograms within the each of the grid cells of the 4×4 sample array used in the SIFT descriptor. For this reason, we call it Histograms of Normalized Colors (HoNC). Since the sample array is translated, rotated, and scaled according to the keypoint location, orientation, and size, the descriptor is invariant to similarity transformations. While it is not invariant to skew and perspective distortion, it is not sensitive to moderate changes in these values. Further invariance can be gained through the use of affine-invariant keypoint detection and representation techniques [19], [20].

In order to gain invariance to illumination intensity and shift, the image colors must be normalized. For each keypoint, we compute the means (μ_r, μ_g, μ_b) and standard deviations (σ_r, σ_g, σ_b) of the color channels in the neighborhood of the keypoint. We then modify the keypoint color values such that the average color value is 127.5 and the average of the three standard deviations is 48:

$$\beta = 127.5 - (\mu_r + \mu_g + \mu_b)/3 \tag{1}$$

$$\gamma = 48 * 3/(\sigma_r + \sigma_g + \sigma_b) \tag{2}$$

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} \gamma(R - \mu_r) + \mu_r + \beta \\ \gamma(G - \mu_g) + \mu_g + \beta \\ \gamma(B - \mu_b) + \mu_b + \beta \end{bmatrix} \tag{3}$$

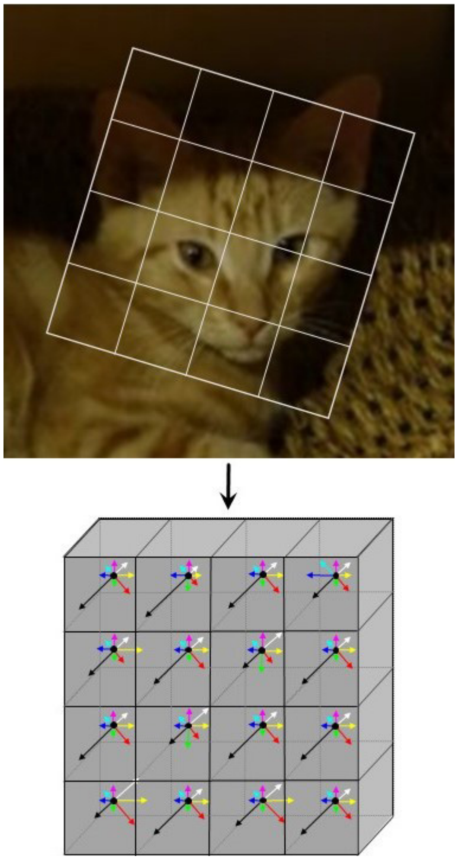


Figure 2. Illustration of the descriptor construction using color histograms. An eight-bin color histogram is formed in each cell of the rotated, scaled, and translated sample array.

After this transformation, the average color mean and standard deviation have been set to a constant value and this yields invariance to illumination intensity and shift. The normalized color values are not invariant to illumination color. Few descriptors are, including the original SIFT descriptor [4]. An exception is RGB-SIFT, which was among the top four descriptors in the study of van de Sande et al. [4], while the other three top descriptors were not. We believe that most applications tolerate variance to illumination color and foregoing it allows better discriminatory power when this is the case. When invariance to illumination color is required, it can be achieved with our descriptor by normalizing each of the color channels with a separate bias and gain:

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} \gamma_r(R-\mu_r)+\mu_r+\beta_r \\ \gamma_g(G-\mu_g)+\mu_g+\beta_g \\ \gamma_b(B-\mu_b)+\mu_b+\beta_b \end{bmatrix} \tag{4}$$

Once the colors have been normalized, we construct color histograms in each of the grid cells of a SIFT-like sample array. In order to compute a descriptor of comparable size, we use a coarse $(2 \times 2 \times 2)$ color histogram in each cell. This yields a $4 \times 4 \times 8 = 128$ dimensional descriptor similar to SIFT. Fig. 2 illustrates the descriptor construction. In each of the cells of the 4×4 image grid (rotated and scaled appropriately) an 8 bin color histogram is constructed corresponding to the primary and secondary colors of light.

As in the SIFT descriptor, we weight the histogram votes using a Gaussian function of the distance from the keypoint center. In each color histogram, interpolation is used not only among four spatial positions, but also among the eight color bins. For each color C, the weights are computed as:

$$w_0 = \min(1, \max(0, (C-63.5)/128)) \tag{5}$$

$$w_1 = 1 - w_0 \tag{6}$$

Appropriate weights for the three color channels are multiplied (and combined with positional weights) for the contribution to each of the histogram bins. This allows a smooth transition between color bins in each of the three color channels. A pixel may contribute to a single bin, if it is near a corner of the color cube, or all eight, if it is near the center.

While this descriptor can be used by itself, we also concatenate it to the SIFT descriptor (or a variation) to form a larger descriptor that captures both color and shape information. Figure 3 shows an example where such a combination was used to achieve improved results over SIFT (compare to Fig 1).

IV. DESCRIPTORS

We compare against (and in combination with) several other descriptors.

A. SIFT

The original Scale-Invariant Feature Transform (SIFT) descriptor was described by Lowe [2]. We use the OpenCV contributed implementation based on the code of Hess [12]. The SIFT descriptor and a variation called GLOH were found to be the best performing local descriptors by Mikolajczyk and Schmid [9] when compared to several grayscale descriptors.

B. SURF

Speeded Up Robust Features (SURF) [14] was designed to be a faster feature detector and descriptor. The feature detector approximates the Hessian determinant using integral images and we use it in our experiments to determine the feature locations for all descriptors. We use the standard descriptor (64 dimensions) in the OpenCV contributed implementation.

C. RGB-SIFT

RGB-SIFT is the concatenation of the SIFT descriptors computed separately for three RGB channels, yielding a 384-dimensional descriptor. This descriptor was considered in the study by van de Sande et al. [4] and found to be among the top performers.

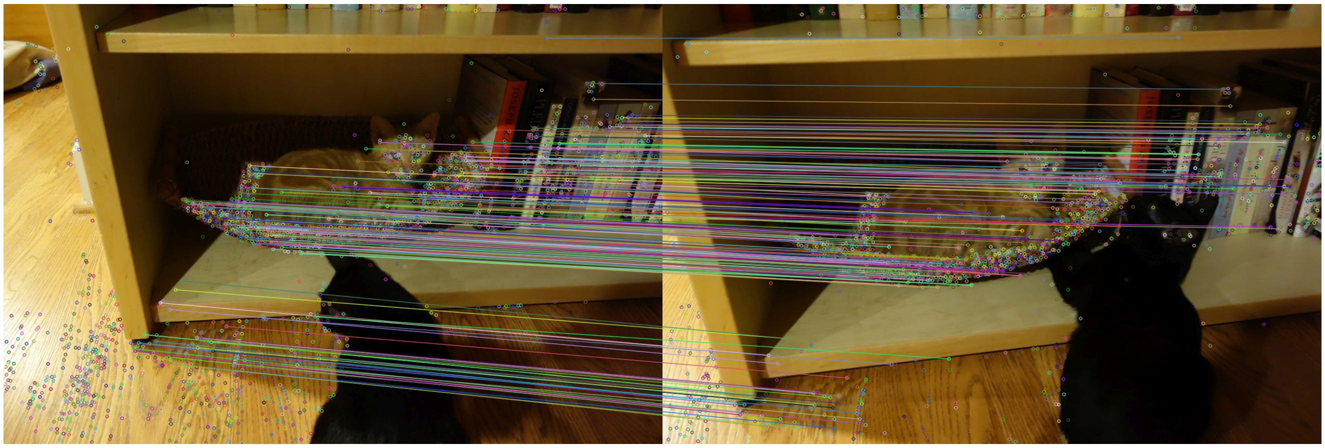


Figure 3. Example of HoNC+SIFT keypoint matching on images of furry creatures. The top 200 matches are displayed. With this descriptor combination, no incorrect matches are found. (Compare to the results using only SIFT in Fig. 1.)

D. rgSIFT

The rgSIFT descriptor [4] concatenates the SIFT descriptors computed on the following two channels, yielding a 256-dimensional descriptor:

$$\left[\begin{array}{c} r \\ g \end{array}\right] = \left[\begin{array}{c} R/(R+G+B) \\ G/(R+G+B) \end{array}\right] \tag{7}$$

E. OpponentSIFT

OpponentSIFT is the same as RBG-SIFT, except that the color channels are first transformed into the opponent color space:

$$\left[\begin{array}{c} O_1 \\ O_2 \\ O_3 \end{array}\right] = \left[\begin{array}{c} (R-G)/\sqrt{2} \\ (R+G-2B)/\sqrt{6} \\ (R+G+B)/\sqrt{3} \end{array}\right] \tag{8}$$

Some previous researchers using this technique have normalized the SIFT descriptors for each channel separately. However, this causes channels with less variation to have a magnified impact on the overall descriptor. We first concatenate the channel descriptors and then perform normalization. Separate normalization was as good (or even better) on many image sets, but was *much* worse on some.

Like RGB-SIFT, this yields a 384-dimensional descriptor and was one of the top performing descriptors in the study by van de Sande et al. [4].

F. C-SIFT

The variation on the C-SIFT descriptor described by van de Sande et al. [4] concatenates the SIFT descriptors for the channels givens by $O1/O3$ and $O2/O3$, yielding a 256 dimensional descriptor. Similar descriptors were proposed by Abdel-Hakim and Farag [17] and Burghouts and Geusebroek [3].

Table I
CHARACTERISTICS OF DESCRIPTORS.

Name	Size	Works on Grayscale	Illumination Invariance
SIFT	128	yes	intensity + shift
SURF	64	yes	intensity + shift
RGB-SIFT	384	yes	intensity + shift + color
rgSIFT	256	no	intensity
OpponentSIFT	384	yes	intensity + shift
C-SIFT	256	no	intensity
HoWH	128	no	intensity + shift
HoNC-none	128	yes	none
HoNC	128	yes	intensity + shift
HoNC-full	128	yes	intensity + shift + color

G. HoWH: Histogram of Weighted Hues

Luke, Keller, and Chamorro-Martinez [6] have suggested stacking the SIFT descriptor with a similar descriptor that replaces the gradient orientation and magnitude at each pixel with the hue and saturation. We consider here a version that is not (necessarily) stacked with the SIFT descriptor. Since hue is a circular value (like gradient orientation) and the saturation describes the strength of the hue (comparable to gradient magnitude), a similarly structured descriptor results. Pixel hue weighted by saturation was also used by van de Weijer and Schmid [5]. However, they use a single histogram for each keypoint that does not combine location and color information. It is also used in combination with a separate shape descriptor.

H. Summary

Table I summarizes the characteristics of the descriptors used in the tests. Included for comparison are two variations on the HoNC descriptor, HoNC-none (in which no color normalization is performed) and HoNC-full (in which colors are normalized separately).

V. Experiments

In order to recognize keypoints between images, we first detect the keypoints in the images using the SURF feature detector [14]. We have found that this yields better matching performance that the SIFT feature detector in our experiments. The top 1000 keypoints are retained for each image. Descriptors for each keypoint are constructed using the techniques described above. Individual descriptors may be used, but we also consider combinations of descriptors that are stacked into longer descriptors. When combined, each individual descriptor vector is scaled to have the same length, regardless of size (except as noted below).

The best match for each keypoint in the reference image is found in the target image using the Euclidean distance between the keypoint descriptors. Matches are considered correct if the projection of the keypoint location into the other image (using a known homography) lies within the computed size of the corresponding keypoint (and in reverse).

We measure the matching performance of each descriptor using the mean average precision as follows. The precision and recall are defined as:

$$\text{precision} = \frac{\text{\# correct matches detected}}{\text{\# total matches detected}} \tag{9}$$

$$\text{recall} = \frac{\text{\# correct matches detected}}{\text{\# keypoints possible to detect}} \tag{10}$$

In computing the recall, we exclude from the denominator those keypoints from the reference image that do not appear in the target image (because they have moved outside the boundaries of the image). We do not exclude keypoints that appear in the target image, but that are missed by the keypoint detector. As the threshold on descriptor distance varies, the number of matches changes and the precision versus recall can be plotted. The *average precision* is the average of the precision over the interval $r \in [0, 1]$ (the area under the curve). The *mean average precision* computes the mean over multiple plots. The maximum value is one and the minimum is zero. Fig. 4 shows an example plot of precision versus recall for one image pair from the Oxford data set.

A. Oxford Data Set

We ran experiments with each descriptor and several combinations on the Oxford affine covariant regions data set[1] that models variations in viewpoint, rotation, zoom, lighting, blur, and compression. All six images (five pairs with the same reference image) of each of the eight data subsets were used. Note that some pairs are quite difficult, with no combination of descriptors achieving an average precision above 0.05.

Table II shows the results. Individually, the top performing descriptors are OpponentSIFT, RGB-SIFT, SIFT, and HoNC.

[1] http://www.robots.ox.ac.uk/~vgg/data/data-aff.html

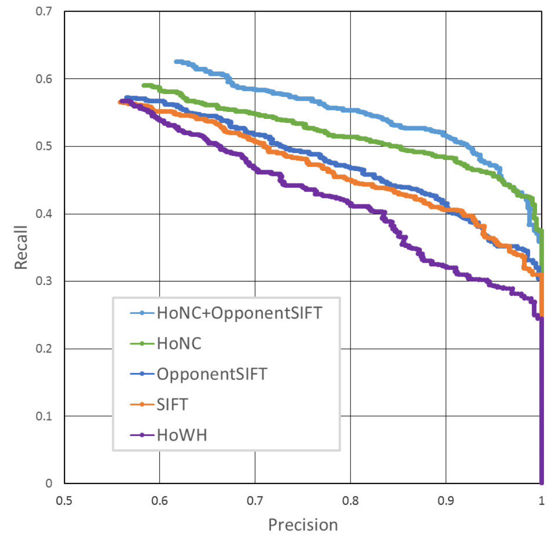


Figure 4. Example precision/recall plot for the first pair of "trees" images in the Oxford data set.

Table II
MEAN AVERAGE PRECISION FOR DESCRIPTORS OVER ALL IMAGE PAIRS IN THE OXFORD DATA SET.

Descriptor(s)	Size	Mean average precision
HoNC+OpponentSIFT	512	.5550
HoNC+RGBSIFT	512	.5535
HoNC+SIFT	256	.5522
HoWH+OpponentSIFT	512	.5421
HoWH+RGBSIFT	512	.5416
HoWH+SIFT	256	.5388
OpponentSIFT+SURF	448	.5250
OpponentSIFT	384	.5237
OpponentSIFT+RGBSIFT	768	.5223
OpponentSIFT+SIFT	512	.5203
RGBSIFT+SURF	448	.5197
RGBSIFT	384	.5185
RGBSIFT+SIFT	512	.5167
SIFT+SURF	192	.5148
SIFT	128	.5136
HoNC+SURF	192	.5074
HoNC+HoWH	256	.5064
HoNC	128	.5043
OpponentSIFT+rgSIFT	640	.4914
OpponentSIFT+CSIFT	640	.4913
SIFT+CSIFT	384	.4911
RGBSIFT+CSIFT	640	.4907
RGBSIFT+rgSIFT	640	.4904
SIFT+rgSIFT	384	.4903
HoNC+CSIFT	384	.4838
HoNC+rgSIFT	384	.4742
SURF+CSIFT	320	.4323
HoWH+rgSIFT	384	.4207
HoWH+CSIFT	384	.4144
SURF+rgSIFT	320	.4036
SURF	64	.4008
rgSIFT+CSIFT	512	.3974
rgSIFT	256	.3940
HoWH+SURF	192	.3939
CSIFT	256	.3900
HoWH	128	.3206

Table III
COMPARISON OF DIFFERENT LEVELS OF NORMALIZATION AND HISTOGRAM COARSENESS WITH THE OXFORD DATA SET. EACH ENTRY IS THE MEAN AVERAGE PRECISION.

Descriptor	Solo	+SIFT	+RGBSIFT	+OpponentSIFT
HoNC-none	.3479	.4608	.4617	.4638
HoNC	.5043	.5522	.5535	.5550
HoNC-full	.5009	.5490	.5502	.5519
HoNC-$3\times3\times3$	.5068	.5613	.5627	.5635

When combinations are used, the three top performers are the combination of HoNC and one of the three top shape-based descriptors. While HoNC+OpponentSIFT was the top combination, HoNC+SIFT performed nearly as well with a descriptor of half the size and will usually be the top choice for this reason. Most of the color information captured by RGB-SIFT and OpponentSIFT is already incorporated into the HoNC descriptor.

When used in combination with other descriptors, the hue-based descriptor (HOWH) performs better when it is shortened (reducing its influence). We used a weight of 0.6, as suggested by van de Weijer and Schmid [5] in their work using a hue-based descriptor. The HoNC descriptor did not require rescaling and performs well over a range of vector lengths, suggesting that it is a more robust descriptor than HoWH. Interestingly, even though HoWH performed poorly by itself, it performed well when combined with a shape-based descriptor.

Unsurprisingly, all of the top performing descriptors combined a color-based descriptor (HoNC or HoWH) with a shape-based descriptor (SIFT, RGB-SIFT, or OpponentSIFT) and these combinations performed significantly better than any individual descriptor. The combination of HoNC with any other descriptor yielded performance that improved upon the additional descriptor by itself. This suggests that any shape-based descriptor could be improved through the use of color information in this manner, even if gradients of color channels are already used as in RGB-SIFT and OpponentSIFT.

SURF, rgSIFT, and C-SIFT were not competitive with the top performing descriptors in these experiments.

B. *Effects of Normalization*

It is interesting to examine the effect of the color normalization by comparing HoNC with HoNC-none (no normalization) and HoNC-full (normalization separately in each color channel, and thus invariant to illumination color). Table III shows a comparison for the Oxford data set. The standard HoNC descriptor performs better than either of the variations, although HoNC-full is close. This supports the observation that full invariance to illumination color is often unnecessary and can be counter-productive.

C. *Finer Histograms*

Needless to say, a $2\times2\times2$ color histogram requires a coarse discretization of the color space. However, finer histograms require additional space and time. We have tested a $3\times3\times3$ color histogram in each grid cell that yields a descriptor with 432 dimensions (roughly comparable to RGB-SIFT and OpponentSIFT). This results in some improvement in the mean average precision (see Table III). However, the improvement is small enough that it is unlikely that it be worth the additional cost in most applications.

D. *Sample Images*

Figures 5 and 6 show additional keypoint matching examples from outdoor scenes. In both cases, HoNC+SIFT significantly improves upon SIFT by itself. For both image pairs, there is a change in viewpoint. Note also the changes in lighting on the castle in Fig. 5 and the changed brightness of the ruins and sky in Fig. 6. Using color normalization, the HoNC detector is robust to intensity changes caused by illumination (and also digital bias and gain modifications).

VI. CONCLUSIONS

We have shown that improved keypoint matching performance can be achieved using normalized color histograms. Our new descriptor (HoNC) combines color and location information in the neighbor of the keypoint in a manner that is invariant to position, scale, orientation, and illumination intensity. It can be made invariant to illumination color, if desired. Improved results were demonstrated on a common data set and example image pairs. Further experiments confirmed that the lack of invariance to illumination color is not a significant drawback on many images.

ACKNOWLEDGMENT

This work was supported, in part, by a Worthington Distinguished Scholar award from the University of Washington Bothell. We are grateful to Nicholas Huebner and James Timmerman who coded previous implementations of some of these ideas.

REFERENCES

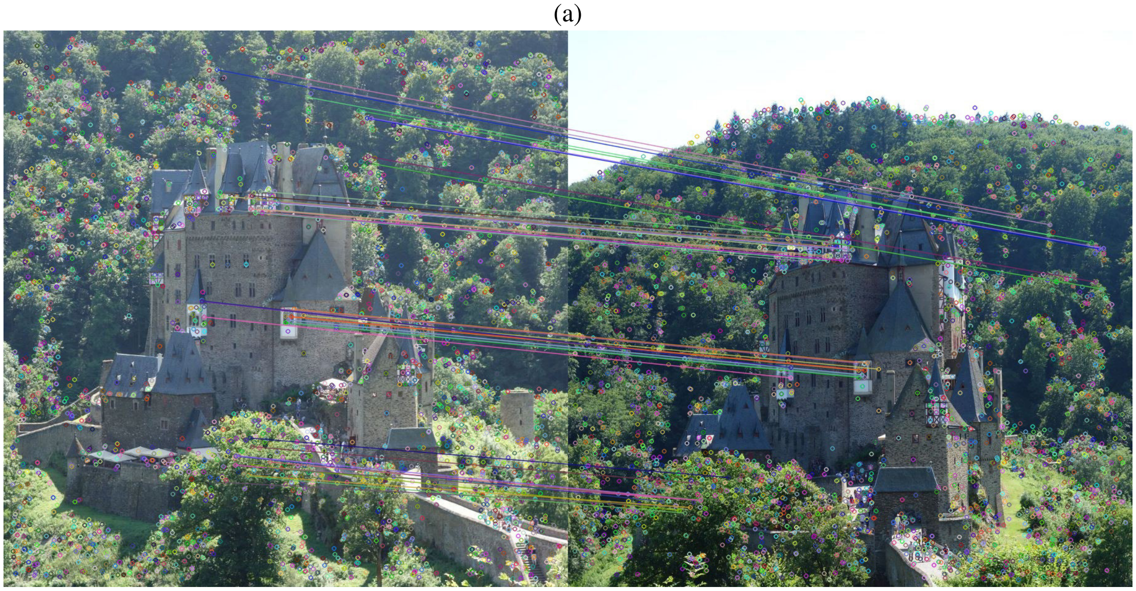
[1] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 5, pp. 530–534, 1997.

[2] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[3] G. J. Burghouts and J.-M. Geusebroek, "Performance evaluation of local colour invariants," *Computer Vision and Image Understanding*, vol. 113, pp. 48–62, 2009.

[4] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek, "Evaluating color descriptors for object and scence recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1582–1596, Sep. 2010.

(a)



(b)

Figure 5. Example of keypoint matching on Eltz Castle. The top 50 matches are displayed. (a) With SIFT matching, eight incorrect matches are found. (b) With HoNC+SIFT matching, no incorrect matches are found.

[5] J. van de Weijer and C. Schmid, "Coloring local feature extraction," in *Proceedings of the European Conference on Computer Vision*, 2006, pp. 334–348.

[6] R. H. Luke, J. M. Keller, and J. Chamorro-Martinez, "Extending the scale invariant feature transform descriptor into the color domain," *ICGST Journal of Graphics, Vision and Image Processing*, vol. 8, no. IV, pp. 35–43, Dec. 2008.

[7] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," *International Journal of Computer Vision*, vol. 73, no. 2, pp. 213–238, 2007.

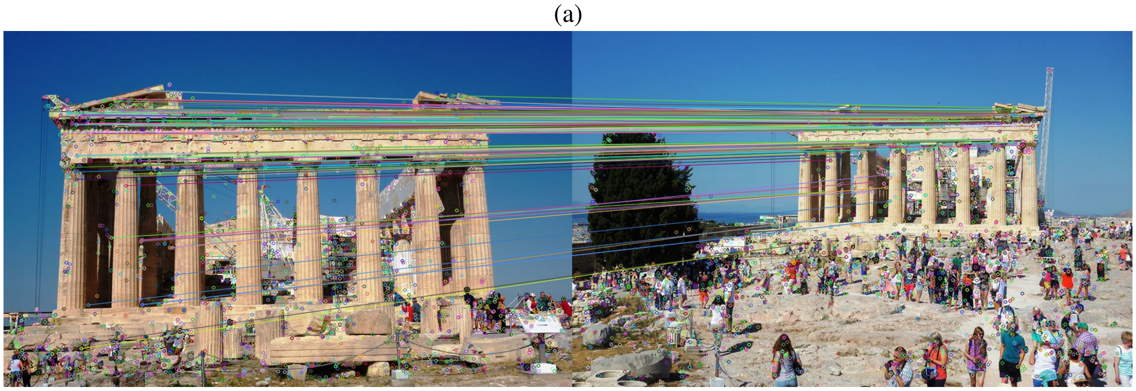
[8] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, no. 1/2, pp. 43–72, 2005.

[9] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, Oct. 2005.

[10] H. Moravec, "Rover visual obstacle avoidance," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 1981, pp. 785–790.

[11] C. J. Harris and M. Stephens, "A combined corner and edge detector," in *Proeedings of the 4th Alvey Vision Conference*, 1988, pp. 147–151.

(a)



(b)

Figure 6. Example of keypoint matching on the Acropolis of Athens. The top 200 matches are displayed. (a) With SIFT matching, there are three qualitatively incorrect matches, in addition to ambiguity between columns. (b) With HoNC+SIFT matching, the matches are largely correct, although ambiguity between columns is present.

[12] R. Hess, "An open-source SIFT library," in *Proceedings of the 18th ACM International Conference on Multimedia*, 2010, pp. 1493–1496.

[13] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 886–893.

[14] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "SURF: Speedup up robust features," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[15] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary robust independent elementary features," in *Proceedings of the European Conference on Computer Vision*, 2010, pp. 778–792.

[16] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proceedings of the International Conference on Computer Vision*, 2011, pp. 2564–2571.

[17] A. E. Abdel-Hakim and A. A. Farag, "CSIFT: A SIFT descriptor with color invariant characteristics," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2006, pp. 1978–1983.

[18] J.-M. Geusebroek, R. van den Boomgaard, A. W. M. Smeulders, and H. Geerts, "Color invaiance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 12, pp. 1338–1350, Dec. 2001.

[19] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.

[20] J.-M. Morel and G. Yu, "ASIFT: A new framework for fully affine invariant image comparison," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 438–469, 2009.