14

14 Color Feature Description

With contributions by Gertjan J. Burghouts

In the previous chapters, we have outlined the theory of invariant feature extraction from color images. The advantage of the full invariants described in Chapter 6 is that they capture intrinsic scenes or object properties, robust to various arbitrary imaging conditions such as local illumination, shadows, and color of the light source. Hence, these invariant features are well suited to characterize the image content in the so-called image descriptors. In this chapter, we demonstrate the appropriateness of such invariant color descriptors. Much of the methodology described here is adopted from Burghouts and Geusebroek [258] and from van de Sande et al. [259].

Many computer vision tasks depend heavily on local feature extraction and matching. Object recognition is a typical case where local information is gathered to obtain evidence for recognition of previously learned objects. Recently, much emphasis has been placed on the detection and recognition of locally (weakly) affine invariant regions [55, 57, 260–262]. The rationale here is that planar regions transform according to well-known laws. Successful methods rely on fixing a local coordinate system to a salient image region, resulting in an ellipse describing local orientation and scale. After transforming the local region to its canonical form, image descriptors should be well able to capture the invariant region appearance. As pointed out by Mikolajczyk and Schmid [252], the detection of elliptic regions varies covariantly with the image (weak perspective) transformation,

Color in Computer Vision: Fundamentals and Applications, First Edition.
Theo Gevers, Arjan Gijsenij, Joost van de Weijer, and Jan-Mark Geusebroek.

while the normalized image pattern they cover and the image descriptors derived from them are typically invariant to the geometric transformation. Recognition performance is further enhanced by designing image descriptors to be photometric invariant such that local intensity transformations due to shading and variation in illumination have no or limited effect on the region description. State-of-the-art methods in object recognition normalize average and standard deviation of the intensity image [55, 252, 263]. Moreover, image measurements using a Gaussian filter and its derivatives are becoming increasingly popular as a way of detecting and characterizing image content in a geometric and photometric invariant way. Gaussian filters have interesting properties from an image processing point of view, among others, their robustness to noise [264], their rotational steerability [265], and their applicability in multiscale settings [54]. Many of the intensity-based descriptors proposed in the literature are based on Gaussian (derivative) measurements [53, 57, 253, 266, 267].

We consider the extension to color-based descriptors as color has high discriminative power. In many cases, objects can well be recognized merely by their color characteristics [43, 46, 47, 268–270]. However, photometric invariance is less trivial to achieve, as the accidental illumination and recording conditions affect the observed colors in a complicated way. Photometric invariance has been intensively studied for color features [46, 47, 50, 58, 164, 187]. The most successful local image descriptor so far is Lowe's SIFT descriptor [55]. The SIFT descriptor encodes the distribution of Gaussian gradients within an image region. The SIFT descriptor is a 128-bin histogram that summarizes local oriented gradients over 8 orientations and over 16 locations. This represents the spatial intensity pattern very well, while being robust to small deformations and localization errors. Nowadays, many modifications and improvements exist, among others, PCA-SIFT [271], GLOH [57], fast approximate SIFT [272], and SURF [273]. These region-based descriptors have achieved a high degree of invariance to overall illumination conditions for planar surfaces. Although designed to retrieve identical object patches, SIFT-like features turn out to be quite successful in bag-of-feature approaches to general scene and object categorization [274].

The important research question is whether color-based descriptors indeed improve on their gray-based counterparts in practice. The answer depends on the stability of the nonlinear combinations of Gaussian derivatives necessary to achieve a similar level of invariance as implemented in gray-value descriptors. For instance, the values of photometric invariants are distorted when the image is JPEG compressed, as the compression distorts the pixel values and spatial layout more for the color channels than for the intensity. Here we provide a study of local color descriptors in comparison with gray-value descriptors.

For (affine) region detection, many well-performing methods exist [149, 252, 254, 275–278]. Hence, we will concentrate on descriptor performance of the full photometric invariant derivatives, as well as their combination into color SIFT descriptors. Furthermore, to enable a fair comparison between intensity-based descriptors and color-based descriptors, we demand identical geometric

invariance for both intensity-based and color-based features. This requirement is conveniently fulfilled by the Gaussian derivative framework.

For the evaluation of local gray-value and color invariants, we adopt the extensive methodology of Mikolajczyk and Schmid [57]. In this article, the authors propose the evaluation of descriptor performance by matching regions from one image to another image. Correct matches are determined using the homography between the two images. From Reference 57, we adopt the measures to evaluate discriminative power and invariance. Also, we adopt variety in recording conditions, being changes of illumination intensity, of the camera viewpoint, blurring of the image, and JPEG compression. We go beyond Reference 57 by extending this set with images recorded under different illumination colors and illumination directions. These conditions induce a significant variation in the image recording. For an illustration of images recorded under varying illumination directions, see Figures 14.1–14.4.

Figure 14.1 An illustration of the diverse objects from the ALOI collection [198]. A random sample of the objects in the collection is depicted.

We extend the number of images used in the evaluation framework [57] to 26,000, representing 1000 objects recorded under 26 imaging conditions. Moreover, we further decompose the evaluation framework in Reference 57 to the level of local gray-value invariants on which common region descriptors are

Figure 14.2 Example object from ALOI recorded under semihemispherical illumination, and images recorded under an illuminant at decreasing altitude angles. See Reference [198] for details.

Figure 14.3 Example object from ALOI recorded under varying illumination color.

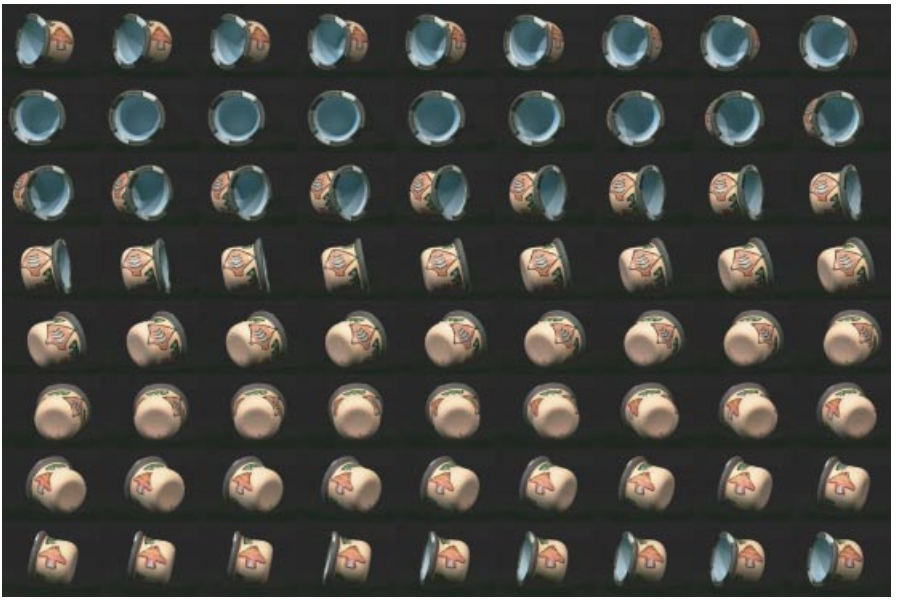


Figure 14.4 Example object from ALOI recorded under varying viewing angles.

based. We measure the performance of photometric invariants for the detection of color transitions only. Hence, we evaluate the performance of the Gaussian gray-value and color invariant derivatives. Finally, we establish performance criteria that are specific to color invariants, indicating the level of invariance with respect to photometric variation, and evaluating the ability to distinguish between various photometric effects.

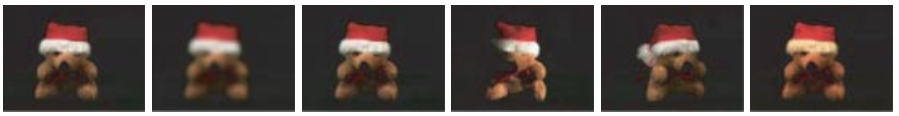
14.1 Gaussian Derivative-Based Descriptors

We compare the local gray-value derivatives with the color invariant derivatives from Chapter 6 based on three evaluation criteria:

- *Discriminative power*. We establish the power of each invariant to discriminate between image regions. Discriminative power is measured by the quality of region matching, similar to Reference 57. The successful matching strategy as proposed by Lowe [55] is based on the rationale that for the recognition of an object it suffices to correctly match only a few regions of that object. In our experimental framework, we push this to the extreme and consider the matching of one region of an object against a database of 1000 regions: one noisy realization of the same object matched against 999 of other objects. Under noisy conditions we consider image deformations caused by blurring, JPEG compression and out-of-plane object rotation (viewpoint change), and photometric variation induced by changes in illumination direction and illumination color. Precision and recall characteristics reflect the discriminative power of the invariant under evaluation.
- *Invariance or robustness.* As above, but now we establish the degradation of the number of correct matches as a function of imaging condition or image transformation that increasingly deteriorates, similar to Reference 279. As with discriminative power, the conditions we test are blurring, JPEG compression, illumination direction, viewpoint change, and illumination color. The degradation in the recall reflects the constancy of the invariant under examination.
- *Information content.* We establish the power of each invariant to discriminate between true color transitions while remaining constant under nonobject-related transitions induced by shadow, shading, and highlights. Hence, we assess simultaneously for each invariant its power to discriminate between color transitions, and its invariance to photometric distortions. Note that this is different from the the two experiments above, as here we evaluate the property to discriminate between the variant and invariant aspects in the photometric condition, in isolation of a possible effect on recognition performance.

We consider for 1000 objects from the ALOI database [67] the following imaging conditions: JPEG Compression; blurring; and changes of the viewpoint, illumination direction, and illumination color. Figure 14.5 illustrates the imaging conditions for some of the objects.

(a)



(b)

Figure 14.5 Randomly selected objects from the ALOI collection (100 example objects) are depicted in (a). Imaging and testing conditions are shown in (b): the reference image, blurring ($\sigma = 2.8$ pixels, image size 192×144), JPEG compression (50%), illumination direction change (to 30° altitude, from the right), viewpoint change (30°), illumination color change (3075K $\rightarrow$ 2175K).

(a) Example image (b) $\bar{E}_w$ (c) $\bar{W}_w$ (d) $\bar{C}_w$ (e) $\bar{H}_w$

Figure 14.6 Photometric invariant gradients. $\overline{E}_w$ is not photometric invariant; $\overline{W}_w$ is invariant to illumination intensity; $\overline{C}_w$ is invariant to shadow and shading; and $\overline{H}_w$ is invariant to shadow, shading, and highlights.

For each object image, we determine its regions. For consistency with the literature, we determine Harris-affine regions [252]. As pointed out in Reference 57, to establish the correct matching of regions, one should either fix the camera viewpoint or consider the homography limiting oneself to more or less flat scenes. For 3D objects, the assertion of a flat scene fails. To overcome this problem, we consider images that have been recorded with a fixed camera viewpoint. However, the condition of viewpoint change has to be settled. Therefore, for each object, we manually selected the single region inside the object that is most consistent between the original and the image recorded under a viewpoint change. We copied the region from the original to all the remaining imaging conditions (see Fig. 14.7 for an example). Note that as we are dealing with regions inside objects only, the black background does not affect the experiments. Furthermore, trying to find 1 region from the 1000 selected regions could be seen as searching the 1 region in an image of 1000 cluttered objects for which all selected regions are visible. Together with the variation in image transformations and imaging conditions a total of 26,000 regions are available. The regions vary significantly in size and anisotropy (Fig. 14.7b and 14.7c). The ground truth of regions is publicly available on the website of the ALOI database [198].

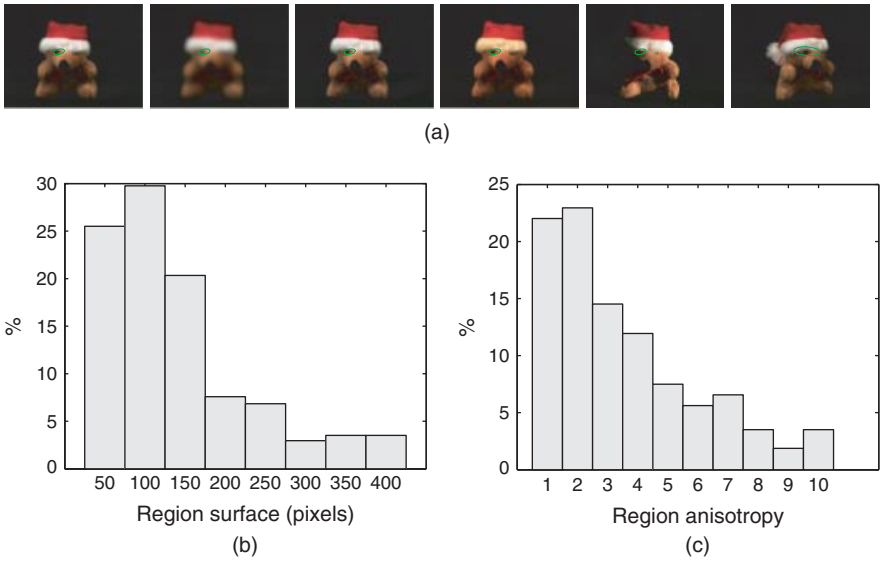


Figure 14.7 (a) Image regions for the reference image, blurring, JPEG compression, illumination color change, illumination direction change, and viewpoint change. For all imaging conditions except the change of viewpoint, the camera is fixed, so the regions are set identical. For the camera viewpoint change, we have manually selected the most stable region. Histogram of (b) the size of the region surfaces, and (c) of the anisotropy (where anisotropy = 1 indicates isotropy).

Next, we compute the invariants from each region. To be consistent with the literature, we normalize the regions as in Reference 252. We consider two experiments:

- *Single location computation* In the first experiment, we compute the invariant gradients from one location. We do so by computing them at a fixed scale (i.e., one-third of the region size). For each region, we determine the location in which the image gradient $\overline{E}_w$ is maximum. For all copied regions (see for region extraction the description above), this location is identical. From this location, we compute all invariants.
- *SIFT-based computation* In the second experiment, we compute the SIFT descriptor from the normalized region identical to Mikolayzcyk's computation [57], but with the gray-value gradient inside the SIFT descriptor replaced by one of the invariant color gradients.

For the performance evaluation, we consider the following sets of invariant gradients (Table 14.1). The extension ''SIFT'' to the name of the invariant implicates SIFT-based computation; otherwise, single-location Gaussian invariants are considered. Original SIFT is also included in the experiments and is equivalent to W-gray-SIFT. We include the intensity gradient W_w in the H and C color-based descriptors. Although this seems contradictory at first sight, the orthogonalization of intensity and intensity-normalized color information proves effective in matching.

Table 14.1 Grey-value and color invariants.

Invariant	Gradients	Property	Equation	Color-SIFT name
E-gray	$\{E_w\}$	Not photometric invariant	—	—
E-color	$\{E_w, E_{\lambda w}, E_{\lambda\lambda w}\}$	Not photometric invariant	6.9	—
W-gray	$\{W_w\}$	Invariant to local intensity level	—	(grey-) SIFT
W-color	$\{W_w, W_{\lambda w}, W_{\lambda\lambda w}\}$	Invariant to local intensity level	6.9	W-color- SIFT
C-color	$\{W_w, C_{\lambda w}, C_{\lambda\lambda w}\}$	Invariant to local intensity level, plus invariant to shadow and shading	6.28	C-color-SIFT
H-color	$\{W_w, H_w\}$	Invariant to local intensity level, plus invariant to shadow and shading, and highlights	6.52	H-color-SIFT

For fair comparison to the original SIFT descriptor, we reduce the dimensionality of all color SIFT descriptors to 128 numbers using PCA reduction (the

covariances have been determined over 200 example regions computed from the reference images). Furthermore, we will evaluate the hue-based SIFT descriptor of Abdel-Hakim and Farag [280], termed *hue-color-SIFT*, and the HSV-based SIFT descriptor of Bosch and Zisserman [281], termed *hsv-color-SIFT*.

14.2 Discriminative Power

The objective of this experiment is to establish the distinctiveness of the invariants. To that end, we match image regions computed from a distorted image to regions computed from the reference images as in Reference 57. The discriminative power is measured by determining the recall of the regions that are to be matched and the precision of the matches:

$$\text{Recall} = \frac{\#\text{correct matches}}{\#\text{correspondences}}, \tag{14.1}$$

$$\text{Precision} = \frac{\#\text{correct matches}}{\#\text{correct matches} + \#\text{ false matches}}. \tag{14.2}$$

Here, recall indicates the number of correctly matched regions relative to the ground truth of corresponding regions in the dataset. Precision indicates the relative amount of correct matches in all the returned matches. The definition of recall is specific to the problem of matching based on a ground truth of one-to-one correspondences, and hence it deviates from the definition as used in information retrieval. The aim in our experiment is to correctly match all regions (recall of one) with ideally no mismatches (precision of one).

We consider the nearest-neighbor matching as employed in Reference 57. Distances between values of photometric invariants are computed from the Mahalanobis distance (the covariances have been determined over 200 examples computed from reference images). Over various thresholds, the number of correct and false matches are evaluated to obtain a recall versus precision curve. A good descriptor would produce a small decay in this curve, reflecting the maintenance of a high precision while matching more image regions.

We randomly draw a test set of regions and use 1000-fold cross-validation to measure performance over our dataset. To end up with graphs that allow a comparison between various levels of color invariance, we vary the number of regions to match per experiment. The number of regions to which a single region is compared is set to 20 for the invariants computed from one location. We consider a successful distinction between 20 image points to be the minimal requirement of a point-based descriptor. For the SIFT-based computation of invariants, we increase this number, as the region-based description is more distinctive. The number of regions to which one region is compared is between 100 or 500, depending on the difficulty of the imaging condition. We consider a successful distinction between 100 regions to be the minimal requirement of a region-based descriptor. We

consider a successful distinction between 500 regions to be sufficient for realistic computer vision tasks; this is in line with the validation in References 57, 279.

The results of the SIFT region matching for invariant gradients are shown in Figure 14.8. All photometric invariants are plotted using solid lines. All color-based invariants are plotted using red lines, opposed to gray-value invariants that are plotted in black lines.

Overall, all color invariants have better performance than gray-value-based features. Gray-value derivatives E-gray and W-gray are outperformed by color-based descriptors, except when illumination color is changed (Fig. 14.8e). In that case, normalized intensity W-gray performs reasonably, but is still outperformed by many color-based invariants, as expected.

The performance of H-color is a bit disappointing compared to the other color invariants. Two effects play a role here. First, this descriptor misses one color channel of information, and better discriminative power could be achieved when adding a saturation channel. However, in that case one would, at best, expect a performance similar to W-color. We will see a comparison later on when establishing performance for the color SIFT descriptors. A second issue affecting the H-color feature is the instabilities caused by the normalization in the denominator of Equation 6.52. The expression becomes unstable for colors that are unsaturated, and hence is grayish. Blurring by the Gaussian filter enhances this effect, as color at boundaries—which we are evaluating in this setup—are mixed. Hence, H-color seems unsuitable for region descriptors based on Gaussian derivatives.

The effect of blurring, shown in Figure 14.8a, causes the image values to be smoothed. Hence, details are lost, but no photometric variation is introduced. The color gradient with no photometric invariant properties, E-color, performs best. Besides the decay in performance due to additional blur, the graph clearly illustrates the gain in discriminative power when using color information.

The compression of images by JPEG, shown in Figure 14.8b, causes the color values to be distorted more than the intensity channel. Still, color information is distinctive, as the color gradient that is invariant to the intensity level, W-color, performs best. At the beginning of the recall-precision curves, one clearly sees the advantage of orthogonalizing intensity and color information, as W-color, C-color, and H-color perform significantly better than E-color, for which all channels are correlated with intensity. In the latter case, all values of the SIFT descriptor will be severely corrupted by the JPEG compression. For the invariant color descriptors, the intensity channel will be relatively mildly corrupted by the compression, whereas the color channels still add extra discriminative power. Compression effects become more influential at the tail of the recall-precision curves, where one sees H-color drop off quite early because of the instability of the descriptor, followed by C-color. Although W-color had a slower start, it ends up doing quite well because of the more stable calculation of the nonlinear derivative combination.

For changes in the illumination direction (Fig. 14.8c), the main imaging effects are darker and lighter image patches, and shadow and shading changes. However,

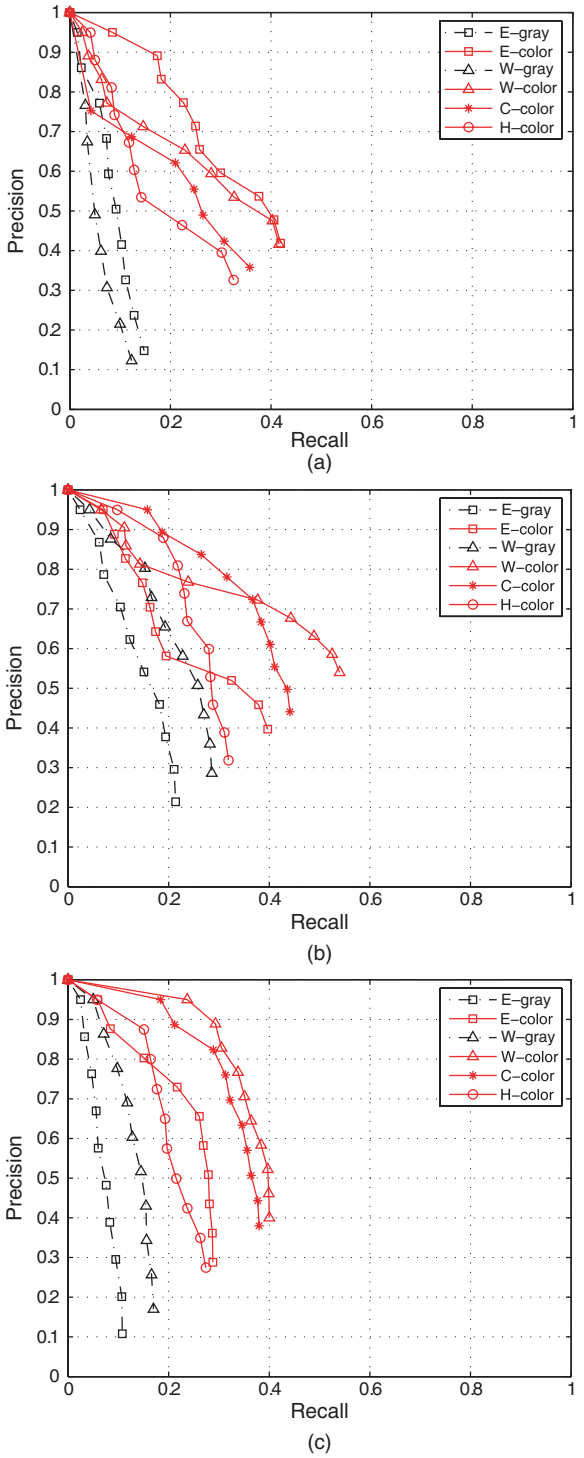
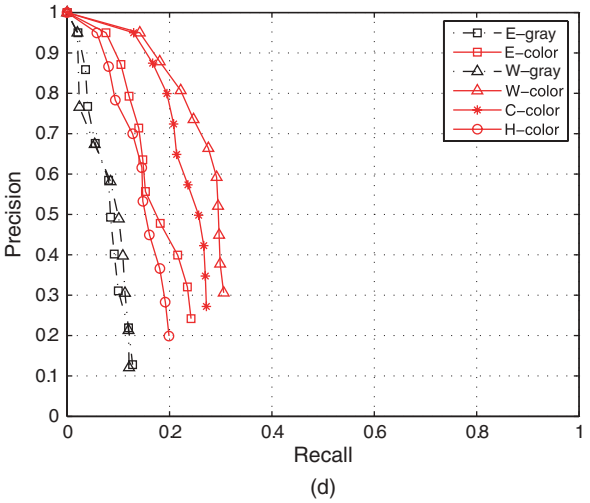
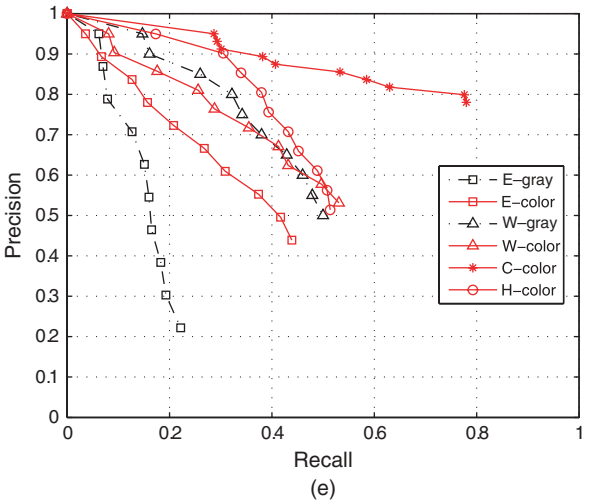


Figure 14.8 Discriminative power of photometric invariant gradients. (a) Blurring ($\sigma = 1$ pixel), 1 versus 20, (b) JPEG compression (50%), 1 versus 20, (c) Illumination direction (30°), 1 versus 20, (d) viewpoint change (30°), 1 versus 20, (e) illumination color (2100K), 1 versus 20.



(d)



(e)

Figure 14.8 (*Continued*)

for the small scale at which we measure the Gaussian derivative descriptors, we expect intensity changes to dominate over shadow and shading edges. Shadow and shading (geometry) edges are expected to become more important when assessing SIFT-based descriptors, which capture information over a much larger region. Hence, both color gradients that are invariant to intensity changes, W-color and C-color, perform well. Clearly, the color invariant descriptors outperform gray-value descriptors and noninvariant color descriptors.

The results of a change in viewpoint (Fig. 14.8d), clearly demonstrate the advantage of adding color information. The patches, manually indicated to be stable, merely contain a change in information content due to a projective transformation and small errors in the affine region detection. Furthermore, the

light field will be distributed somewhat differently over the image, causing W-color and C-color to perform superior to gray-value descriptors, noninvariant color descriptors, and the H-color descriptor.

For varying illumination color (Fig. 14.8e), obviously the color values become distorted. The color gradient invariant to shadow, C-color, is very robust here. Although C-color is based on color, its gradients are computed in such a way that can be shown to be reasonably color constant [67]. Furthermore, one would expect the gray-value descriptors not to be affected by illumination color changes. However, a change in overall intensity is also present, making direct use of E-gray infeasible. The intensity-normalized invariant W-gray performs reasonably, but lacks the discriminative power that comes with the use of color.

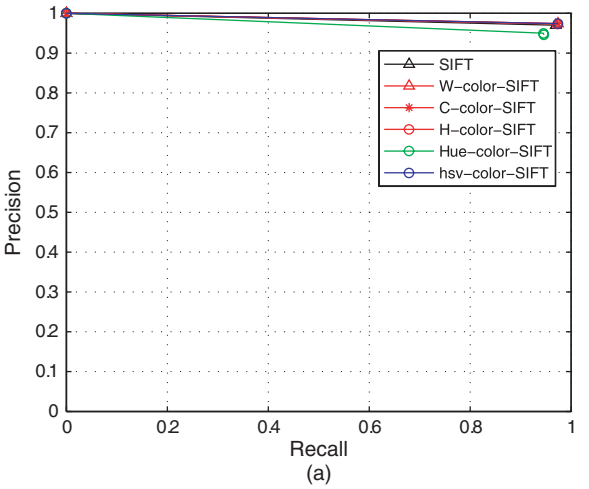
Figure 14.9 shows the discriminative power of the invariants when they are plugged into the SIFT descriptor. The figure has an identical organization as Figure 14.8. The only exception in the experimental setup is that the number of regions, to which a single region is matched, is increased. This number varies over the imaging conditions, and is either 100 or 500, to obtain suitable resolution in the performance graphs. Furthermore, note that two extra methods from the literature have been added, the hue-color-SIFT descriptor [280] and the hsv-color-SIFT descriptor [281].

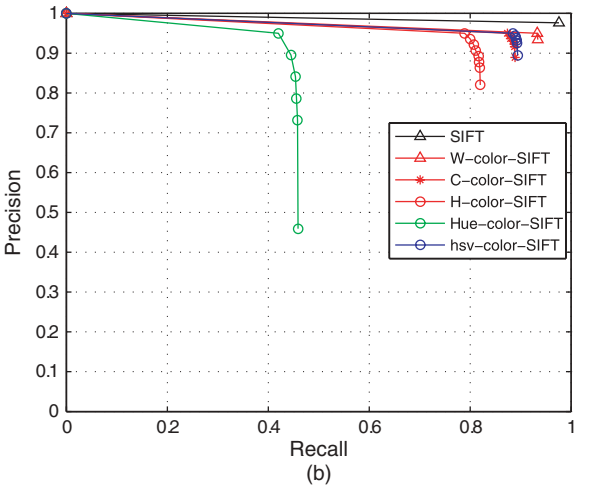
Overall, the relative performance of SIFT-based computation of invariants corresponds largely to the relative performance of invariants from single points. Color-based SIFT invariant to shadow and shading effects, C-color-SIFT, performs best.

Generally, the SIFT-based computation significantly improves the discriminative power compared to single-point computation. Almost all color and gray-value descriptors perform well under blurring (Fig. 14.9a), JPEG compression (Fig. 14.9b), and illumination color changes (Fig. 14.9e). Note that the C-color-SIFT descriptor performs equally well as the intensity-based SIFT descriptor in the case of illumination color changes, implying a high degree of color constancy for this descriptor.

Discriminative power drops when considering illumination direction or viewpoint changes (Fig. 14.9a,b). These cases are much harder to distinguish using a SIFT descriptor. In these cases, the gray-value-based SIFT is outperformed by the color-based SIFT descriptors. In particular, the color-based SIFT invariant to shadow and shading effects, C-color-SIFT, is very discriminative in these cases. This can be explained by the large spatial area over which the SIFT descriptor captures image structure. Hence, shadow and shading (object geometry) effects are more likely to be captured by the SIFT descriptor, but the effects are cancelled by the C invariant.

The shadow and highlight invariant H-color-SIFT is generally not very distinctive compared to W-color-SIFT and C-color-SIFT. Lack of discriminative power affects the performance of hue-color-SIFT, H-color-SIFT, and SIFT under blurring. Furthermore, the hue-based descriptors, hue-color-SIFT, and H-color-SIFT are affected by JPEG compression and by illumination color changes. The distinctiveness of hue-color-SIFT is generally much less than that of H-color-SIFT.





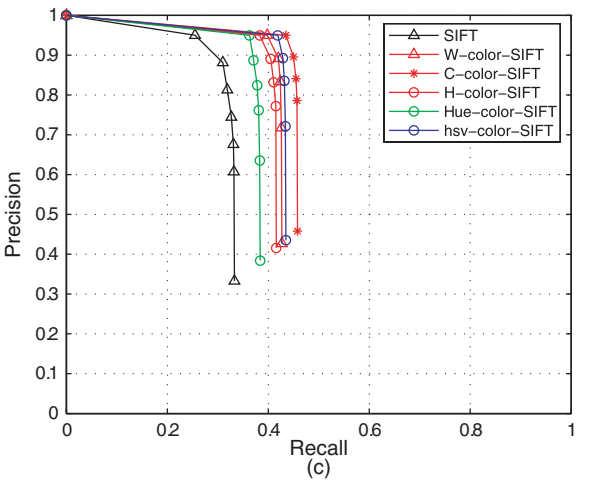
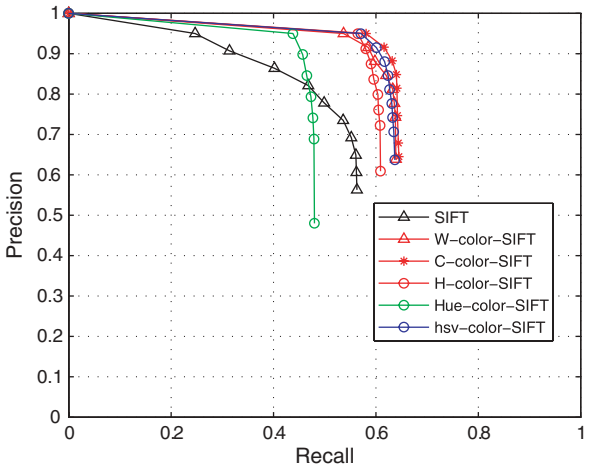
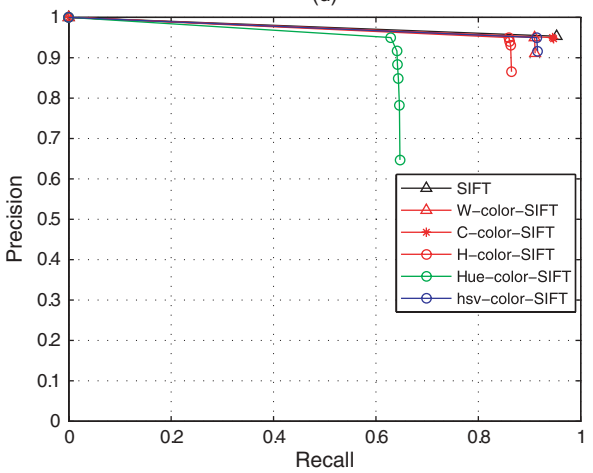


Figure 14.9 Discriminative power of photometric invariant gradients when plugged into the SIFT descriptor. (a) Blurring ($\sigma = 1$ pixel), 1 versus 500, (b) JPEG compression (50%), 1 versus 500, (c) Illumination direction (30°), 1 versus 100, (d) Viewpoint change (30°), 1 versus 100, (e) Illumination color (2100K), 1 versus 500.



(d)



(e)

Figure 14.9 (*Continued*)

Hence, using the hue alone is not a distinctive region property. The distinctiveness of hsv-color-SIFT is generally somewhat higher than that of H-color-SIFT. Thus, the saturation s in the *hsv* color space is a distinctive property. However, the distinctiveness of hsv-color-SIFT is generally less than that of W-color-SIFT and C-color-SIFT because of Instability, as argued before.

14.3 Level of Invariance

The objective of this experiment is to establish the constancy of the invariants against varying imaging conditions. Likewise [279], we measure the degradation

of recall (Eq. 14.1) over increasingly hard imaging conditions. The experimental setup is identical to that in the previous experiment.

The results of the region matching over increasingly hard imaging conditions is shown in Figure 14.10. The organization of the figure is identical to Figures 14.8 and 14.9. The presented graphs are orthogonal to Figures 14.8 and 14.9, in that now the amount of degradation is varied, at a fixed recall that corresponds to the endpoint of the curves in Figures 14.8 and 14.9. Any decline in performance indicates lack of constancy with respect to the tested condition. Ideally, the decline would be zero (horizontal line), indicating perfect invariance to the set of imaging conditions.

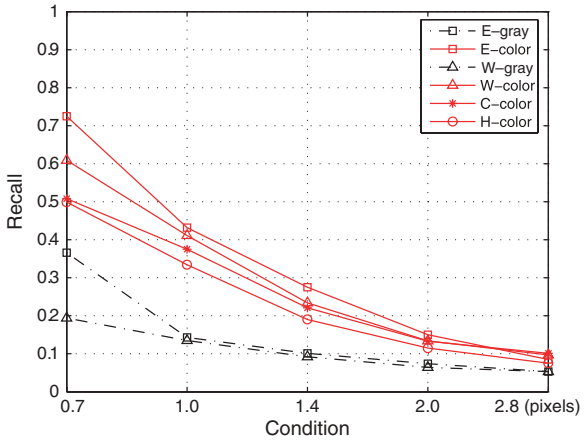
For image blurring (Fig. 14.10a) no significant imaging effects are observed. Hence, all descriptors have equal performance with respect to constancy, although initial discriminative power varies from a recall of 0.2 for gray-value derivatives to more than 0.7 for color-based derivatives. For JPEG compression (Fig. 14.10b), the gray-value invariants E-gray and W-gray are slightly more constant than the color invariants, as the image intensity is less affected by JPEG compression than the image chromaticity. For changes in the illumination direction (Fig. 14.10c) due to the small scale of the derivative descriptors, the main imaging effect is the change in region intensity. Hence, W-gray, W-color, C-color, and H-color are very stable. For a viewpoint change (Fig. 14.10d), only marginal imaging effects are observed. Hence, all measures perform equally well with respect to constancy. For varying illumination color (Fig. 14.10e), besides the intensity-based measures E-gray and W-gray, C-color is very invariant. This measure has theoretically been shown to be reasonably color constant [67].

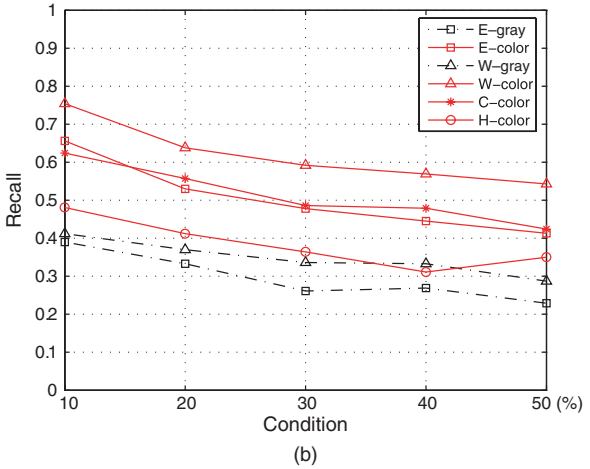
We repeat the invariance experiment, but now the invariants are plugged into the SIFT descriptor. The results are shown in Figure 14.11.

Overall, most descriptors have performed well for blurring (Fig. 14.11a), JPEG compression (Fig. 14.11b), and illumination color change (Fig.14.11e). Exceptions again are the hue-based descriptors H-color-SIFT and hue-color-SIFT, which lack discriminative power, and are more affected by these conditions. A change in illumination direction or viewpoint is much harder for the SIFT descriptor to deal with, even with the color invariance built in. Overall, the C-color-SIFT seems the best choice, for which shadow and shading edges are discounted. This descriptor has invariance comparable to the intensity-based SIFT descriptor, but gains considerably in discriminative power.

14.4 Information Content

The objective of this final experiment is to establish the information content of the photometric invariants. Information content refers to the ability of an invariant to distinguish between color transitions and photometric events such as shadow, shading, and highlights. Ideally, the invariant's values covary with color transitions and its value is constant to photometric events to which it is designed





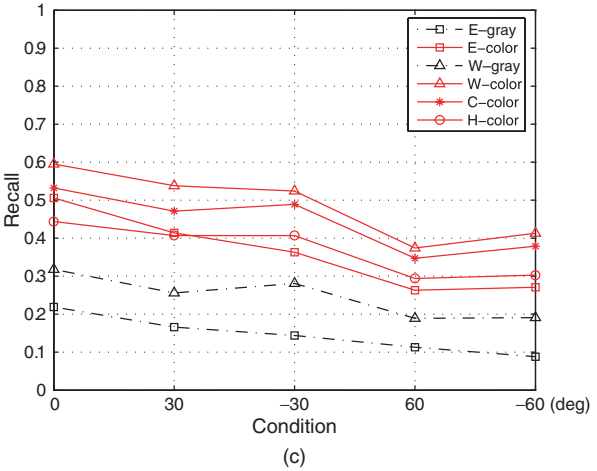


Figure 14.10 Invariance of photometric invariant gradients over increasingly hard imaging conditions. (a) Blurring, 1 versus 20, (b) JPEG compression, 1 versus 20, (c) Illumination direction, 1 versus 20, (d) Viewpoint change, 1 versus 20, (e)Illumination color, 1 versus 20.

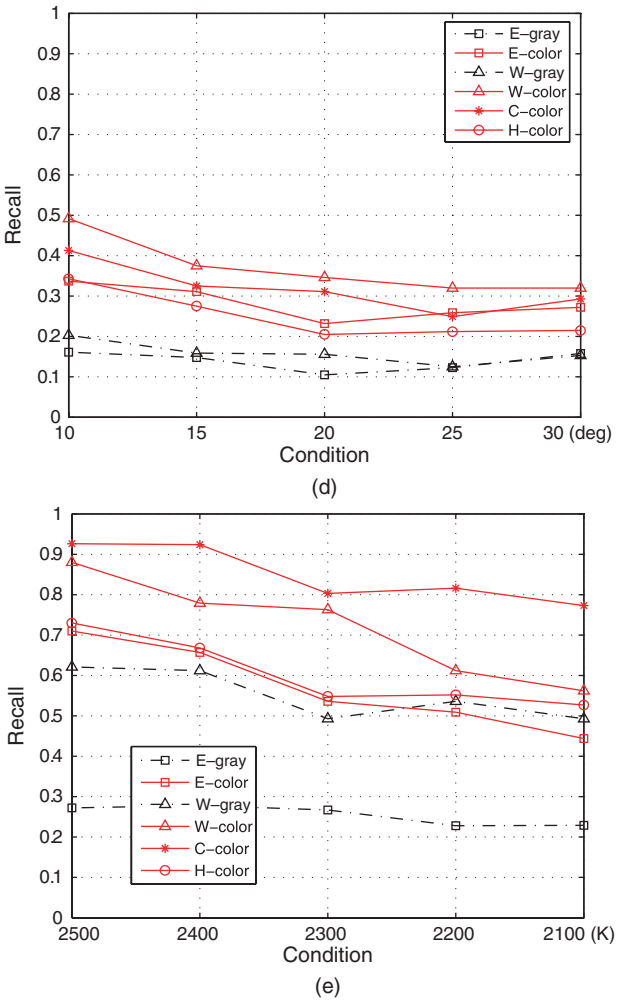


Figure 14.10 (*Continued*)

to be invariant. We illustrate the information content of W-color and C-color (Fig. 14.12). For the first object, new image edges are introduced by changing the illumination direction in Figures 14.12b and 14.12c. Hence, the matching is better with the shadow and shading invariant descriptor C-color-SIFT. Figures 14.12e and 14.12f show an example where no shadow/shading invariance performs better. Here, no new edges are introduced by the change in illumination direction, and only the local intensity is affected because of the relatively large-scale shading effects.

To establish the information content, we measure the discriminative power and invariance over individual image regions. Each image region is labeled whether it contains a color transition, or a shadow, shading, or highlight transition. In this way, the information content evaluates the invariant's discriminative power and

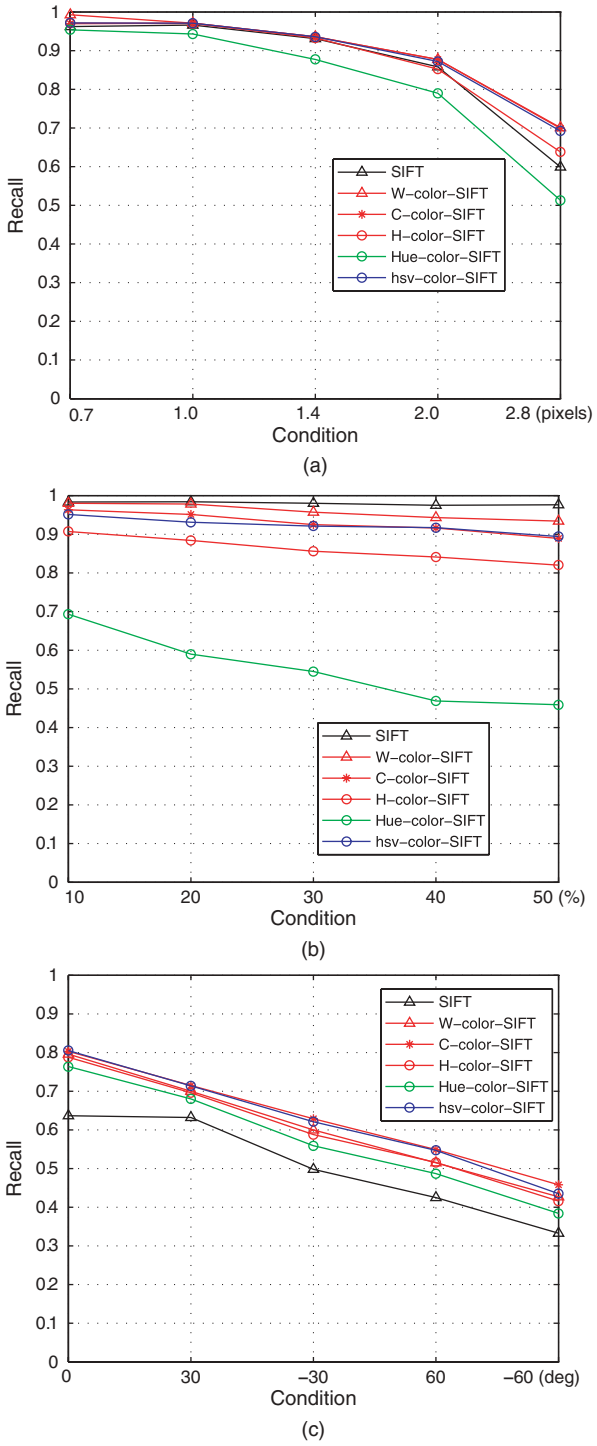


Figure 14.11 Invariance of photometric invariant gradients over increasingly hard imaging conditions when plugged into the SIFT descriptor. (a) Blurring, 1 versus 500, (b) JPEG compression, 1 versus 500, (c) Illumination direction, 1 versus 100, (d) Viewpoint change, 1 versus 100, (e) Illumination color, 1 versus 500.

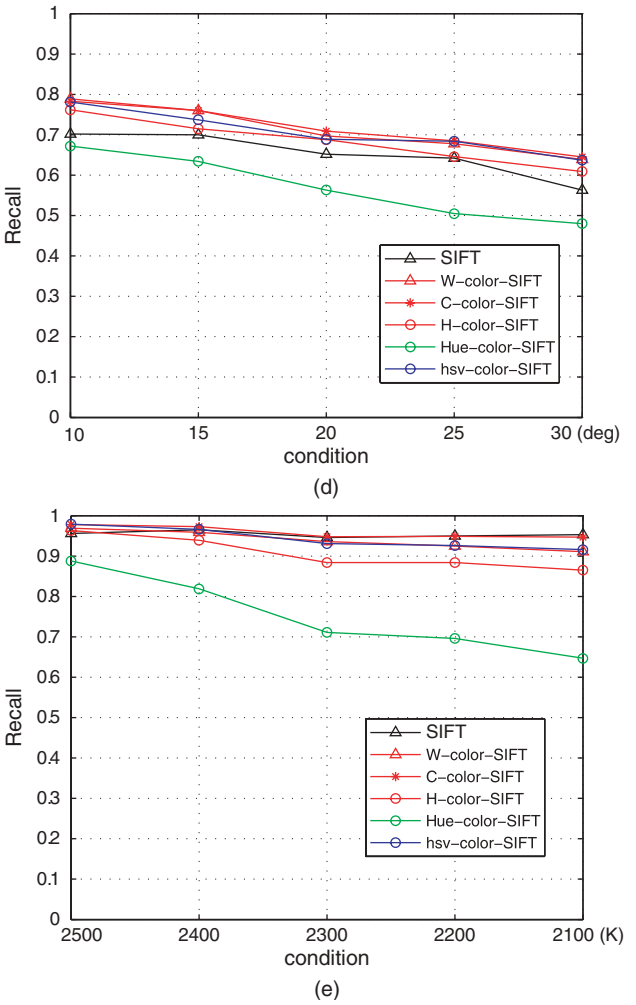
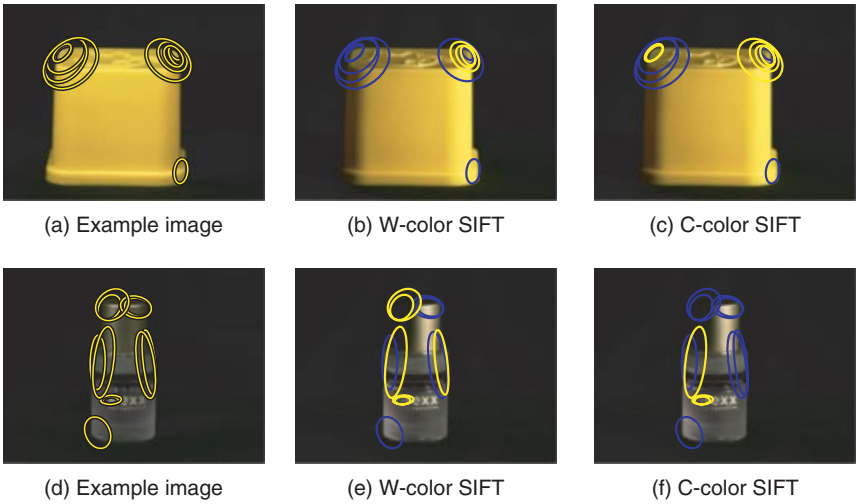


Figure 14.11 (*Continued*)

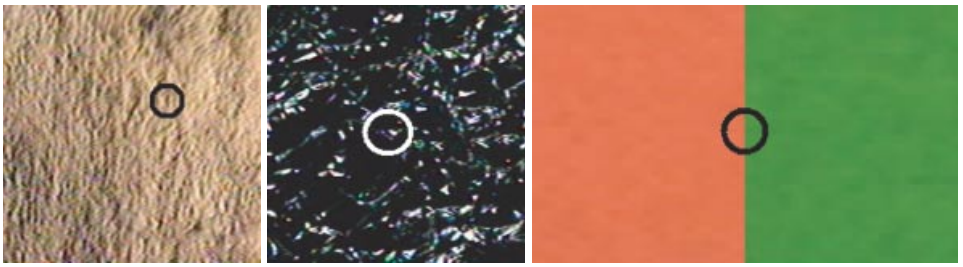
invariance over various photometric events. To that end, we construct a large annotated dataset from images selected from the CURET dataset [282]. This dataset contains tens of images within the order of hundreds of labeled image points located at the various photometric events. The selected texture images contain many edges, where we annotated for each image whether the texture was generated mainly by either shadow/shading (*sponge, cracker b, lambswool, quarry tile, wood b*, and *rabbit fur*) or highlight effects (*aluminum foil, rug a*, and *styrofoam*). From these images, regions have been detected by applying a Harris corner detector [53]. Figure 14.13a,b illustrates, for two fragments of texture images, shadow/shading and highlight edges, respectively. In addition, we have



(a) Example image (b) W-color SIFT (c) C-color SIFT

(d) Example image (e) W-color SIFT (f) C-color SIFT

Figure 14.12 Illustration of matching for two objects. One is better matched with C-color-SIFT, the other with W-color-SIFT, respectively. Correct matches are shown in yellow, false matches are shown in blue. *Source*: Reprinted with permission, © 2009 Elsevier.



(a) (b) (c)

Figure 14.13 Examples of the photometric events dataset. Detected points are given a label whether the point is located on a (a) shadow/shading edge, (b) highlight edge, or (c) color edge.

collected image points located at color transitions. To that end, images have been taken from PANTONE color patches [70] (see Fig. 14.13c for an illustration). From the PANTONE patch combinations, we have selected the 100 combinations that have the largest hue difference, that is, patches that reflect true changes in object color rather than intensity or saturation differences.

We measure an invariant's power to distinguish between color transitions and disturbing photometric events by the Fisher criterion. From many color transitions, we compute a first cloud of points; from transitions of a particular disturbing photometric event, we compute a second point cloud. The Fisher

criterion expresses the separation between the two clouds of points, termed $\{x_1\}$ and $\{x_2\}$ respectively:

$$\text{Information} = \frac{|\mu(\{x_1\}) - \mu(\{x_2\})|^2}{\sigma^2(\{x_1\}) + \sigma^2(\{x_2\})}. \tag{14.3}$$

14.4.1 Experimental Results

The values of photometric invariants to various photometric events are shown in Figure 14.14. The plots show values relative to the total color edge strength $\overline{W}_w$. We do so to express simultaneously the power of $\overline{W}_w$ and of the shadow and shading invariants $\overline{C}_w$ and $\overline{H}_w$ to distinguish between photometric events and true color edges. As expected, the values of the invariants $\overline{C}_w$ and $\overline{H}_w$ are close to zero for shadow/shading edges (note that values of the reference invariant $\overline{W}_w$ are indeed significant to shadow/shading edges). For shadow/shading disturbances, we obtain information$(\overline{C}_w) = 2.6$, and information$(\overline{H}_w) = 4.9$. Thus, the invariant $\overline{H}_w$ separates shadow/shading from object transitions much better than $\overline{C}_w$. Furthermore, the value of $\overline{H}_w$ is also low for highlights (Fig. 14.14b). However, as expected, not all of the values are close to zero because of pixel saturations at highlights. As a result, the invariance and the information content of $\overline{H}_w$ are

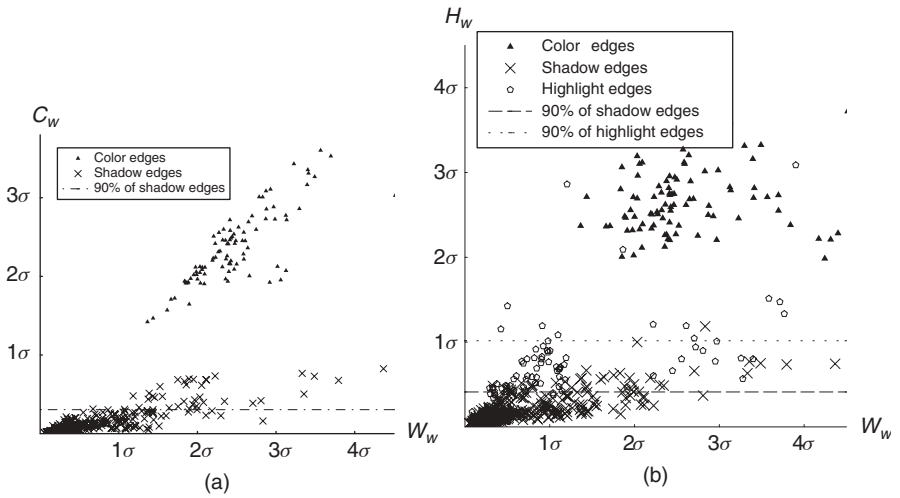


Figure 14.14 Scatter plots of invariant values to photometric events. The figures depict (a) $\overline{C}_w$ versus $\overline{W}_w$ and (b) $\overline{H}_w$ versus $\overline{W}_w$. All invariants are sensitive to color edges. $\overline{C}_w$ and $\overline{H}_w$ are invariant to shadow and shading, where $\overline{H}_w$ is additionally invariant to highlights. The horizontal lines describe a 90% interval of the invariant values. This gives an indication of the invariant's ability to distinguish between values to color edges and to disturbing photometric events. *Source*: Reprinted with permission, © 2009 Elsevier.

somewhat lower for highlight disturbances than for shadow/shading disturbances, $information(\overline{H}_w) = 2.9$.

Overall, the photometric invariant H-color is more constant to shadow and shading than C-color. Both perform well when separating color transitions from shadow and shading transitions. The separation of color transitions and highlights by H-color is harder because of saturated highlights. As a consequence, most of the highlights are separated well, but some highlights are misclassified as color transitions.

14.5 Summary

In this chapter, we have discussed color invariant descriptors for image description and recognition. We evaluated the descriptive power of local derivative-based color invariants and the descriptive power of color invariant SIFT descriptors. In Chapter 16 we will show the application of these descriptors in image and video retrieval.