Evaluation of Low-Level Image Representations for Illumination-Insensitive Recognition of Textureless Objects

Sebastian Zambanini and Martin Kampel*

Computer Vision Lab, Vienna University of Technology, Austria

Abstract. In this paper the problem of recognizing textureless objects in unconstrained illumination and material conditions is investigated. We evaluate the discriminative power of various low-level image features for a pixelwise representation of the underlying surface characteristics of the object. For this purpose, a new dataset with rendered images of 3D models is used which allows to directly compare the influences of texture and material properties in an object recognition scenario. The results are further validated on a dataset of real object images and finally reveal that jets of single- and multi-scale even Gabor filter responses outperform other proposed features in scenarios with textureless objects and strong variations of illumination.

1 Introduction

Achieving invariance to illumination conditions is a major problem of many computer vision tasks. It has been heavily researched in the past in research areas like face recognition [11] or object recognition [14]. However, usually methods presented in these areas have the goal to extract the albedo information (a.k.a. reflectance or intrinsic image) and reduce the illumination effects. In the standard model [2], the image $I(x, y)$ is considered to be the product of the reflectance $R(x, y)$ (i.e. the albedo or texture) and the illumination effects $L(x, y)$, $I(x, y) = R(x, y) \cdot L(x, y)$. For textured objects, this decomposition makes sense because the albedo image $R(x, y)$ provides a discriminative basis for object comparison. For instance, for face images $R(x, y)$ describes the position and shapes of the lips, eyes, eyebrows etc. However, there is also a wide range of objects in the world with constant albedo (i.e. textureless objects), like coins, statues or facades. As for these objects $R(x, y)$ is constant over the entire image, such a decomposition does not provide any new information helpful for object recognition. In this paper we especially focus on this kind of objects. We address the problem of determining if *two aligned images of textureless objects or object parts show the same 3D surface*. We thereby restrict our study to low-level features in scenarios with *arbitrary, unknown illumination conditions* and *arbitrary, unknown*

* This research was supported by the Austrian Science Fund (FWF) under the grant TRP140-N23-2010 (ILAC).

A. Petrosino (Ed.): ICIAP 2013, Part I, LNCS 8156, pp. 71–80, 2013.

bidirectional reflectance distribution function (BRDF) of the objects' material. Furthermore, we do not consider methods which exploit any other additional information, like 3D object models [3] or learned object appearance from training images [9].

Related Work. In fact, there is only few research related to this problem. Local image features for multi-view object matching have been evaluated with regard to their performance under lighting variations [14,16], but these studies were more general and did not especially evaluate the pixelwise low-level representations on textureless objects. Illumination-invariance in general has also been investigated by Chen et al. [5] concluding that without a priori information for general object classes a trueillumination invariant does not exist. Nevertheless, there are representations which are less sensitive to illumination conditions than the original images. Chen et al. proposed to use image gradient directions as they are invariant to linear brightness transformations. Thus, gradient directions are a good choice for flat, textured objects or objects with surfaces of anisotropic depth (i.e. surfaces where the depth changes rapidly in one direction and slowly in another). Another study focusing on low-level features was performed by Osadchy et al. [15]. They showed that the decorrelation induced by a whitening filter for isotropic surfaces increases the distinctiveness of object images. As an approximation for whitening, the Laplacian of Gaussian (LOG) filter was proposed. LOG could be effectively combined with the gradient orientation to a jet of oriented second derivative filters for a distinctive representation for both isotropic and anisotropic surfaces.

Contribution. The problem with these existing studies is that they do not explicitly separate the cases of textured and textureless objects and thus can not give a well-founded statement about the performance of the investigated representations on textureless objects. It is also unclear how the performances of low-level features are related to the material properties or the amount of illumination changes. To explore these questions, we evaluate several low-level representations proposed by earlier works with respect to their performance on textureless and textured objects. This is achieved by a comprehensive evaluation on a new database of synthetically generated images allowing to directly investigate the effects of different material BRDFs and the texturedness of objects. Furthermore, we validate our results on real images of textureless objects. Thus, this paper helps to assess the discriminability and illumination-insensitivity of low-level representations which is helpful for researchers developing superior features and methods for textureless objects. The investigated representations are the basis for more sophisticated local features, e.g. gradient direction [13], gradient orientation [6] or steerable filter responses [4]. Therefore, one can also derive from the presented evaluation how these features act in scenarios with textureless objects.

2 Low-Level Image Representations

We compare eight image representations that have been chosen from literature:

Gradient Direction (GD): Image gradient directions have been identified by Chen et al. [5] as an illumination-insensitive image feature. In the circular domain, the distance between two gradient directions $\mathrm{GD}(p)$ and $\mathrm{GD}'(p)$ at the image point $p = (x, y)$ is computed as the minimum between $(\big|\mathrm{GD}(p) - \mathrm{GD}'(p)\big|)$ and $(2\pi - \big|\mathrm{GD}(p) - \mathrm{GD}'(p)\big|)$. Using this distance metric for the individual pixels, we take the Sum of Squared Differences (SSD) to compare two images.

Gradient Orientation (GO): Instead of representing image gradients in a *signed* version (directions between 0-360 degree), we can also use an *unsigned* version (orientations between 0-180 degree) of gradients. Gradient orientations are in theory less sensitive to the lighting directions than gradient directions, as opposite lighting directions tend to produce opposite gradient directions at depth discontinuities on the surface [15]. From the gradient direction $\mathrm{GD}(p)$, the gradient orientation $\mathrm{GO}(p)$ can be simply computed as $\mathrm{GO}(p) = \mathrm{mod}(\mathrm{GD}(p), \pi)$. To compare two images, the SSD is used where the pixel difference is defined as the minimum between $(\big|\mathrm{GO}(p) - \mathrm{GO}'(p)\big|)$ and $(\pi - \big|\mathrm{GO}(p) - \mathrm{GO}'(p)\big|)$.

Laplacian of Gaussian (LOG): The Laplacian of Gaussian is an approximation of the whitening filter tending to decorrelate the images which makes the filter appropriate for isotropic surfaces [15]. We use the LOG filter by convolving the image and normalizing the absolute responses to unit length. The distance between two images is then again determined by the SSD.

Jets of Gabor Filter Responses (JG): Gabor filters refer to the work of Dennis Gabor [8] in which he proposed to represent a signal as a combination of elementary functions. Gabor filters are widely mentioned to be insensitive against illumination conditions [1,15,12] due to their invariance against additive and multiplicative intensity changes which makes them a popular low-level feature for applications like face recognition [1]. A Gabor filter G has complex coefficients and can thus be defined in terms of a real/even part G_e and an imaginary/odd part G_o,

$$G_e(x,y) = e^{-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}} \cos\left(2\pi\frac{x'}{\omega}\right), G_o(x,y) = e^{-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}} \sin\left(2\pi\frac{x'}{\omega}\right) \quad (1)$$

with $x' = x\cos\theta + y\sin\theta$ and $y' = -x\sin\theta + y\cos\theta$. The parameter σ defines the standard deviation of the Gaussian envelope whereas ω represents the wavelength of the sinusoidal plane wave. To construct Gabor filters of different sizes but equal shapes one can define σ as a linear function of ω, $\sigma = c \cdot \omega$. θ defines the orientation of the filter and γ is the spatial aspect ratio. To construct a Gabor filter bank we keep the parameters σ, c and γ fixed, use N equally spaced orientations $\theta_1 \ldots \theta_N$ and filter the image with the corresponding N Gabor filters $G_e^{\theta_i}$ and $G_o^{\theta_i}$. The jet $\widetilde{\mathrm{JG}}(p)$ is a vector of the magnitude responses of the filtered images $I_e^{\theta_i} = I \star G_e^{\theta_i}$ and $I_o^{\theta_i} = I \star G_o^{\theta_i}$,

$$\widetilde{\mathrm{JG}}(p) = [\sqrt{(I_e^{\theta_1}(p))^2 + (I_o^{\theta_1}(p))^2}, \ldots, \sqrt{(I_e^{\theta_N}(p))^2 + (I_o^{\theta_N}(p))^2}] \quad (2)$$

In addition to complex shading patterns, illumination variations can also induce simple multiplicative changes of image intensities which can be compensated by

normalizing the jet to unit length [12,15]. The final feature is thus given by the normalized jet $\mathrm{JG}(p)$ and the distance between two jets $\mathrm{JG}(p)$ and $\mathrm{JG}'(p)$ is computed as the L2-norm of their vector difference. Image distances are computed by taking the SSD of $\mathrm{JG}(p)$ and $\mathrm{JG}'(p)$ for all image points p.

Jets of Even Gabor Filter Responses (JEG): Besides Gabor jets, jets of oriented second derivatives of Gaussians [7] have also been proposed as an effective way of combining LOG and GO to produce a representation which is appropriate for both isotropic and anisotropic surfaces [15]. Even Gabor filters have a very similar shape to oriented second derivatives of Gaussians if the cosine bandwidth is chosen such that the Gaussian envelope roughly covers the cosine range of $[-1.5\pi, 1.5\pi]$ (i.e., $c \approx 0.4$) [12,15]. In this study we use even Gabor filters as they provide a higher flexibility in the definition of the filter shape, due to a more general set of parameters. However, it is clear from the high similarity of the filters that substantially the same performance can be achieved by the use of second derivatives of Gaussians. In contrast to JG, the jet $\widetilde{\mathrm{JEG}}$ is formed only from the absolute values of $I_e^{\theta_i}$,

$$\widetilde{\mathrm{JEG}}(p) = [\left|I_e^{\theta_1}(p)\right|, \ldots, \left|I_e^{\theta_N}(p)\right|] \tag{3}$$

The final feature is again given by the normalized jet $\mathrm{JEG}(p)$.

Jets of Multi-Scale Even Gabor Filter Responses (JMSEG): The optimal size of the filters depends on the surface characteristics, as for smoother surface parts a wider filter is needed than for less smooth surface parts [15]. One can learn the optimal filter size for a given application domain by means of training data as done in [15], but nonetheless the variation of surface smoothness is disregarded if only one single filter size is used. As in a general scenario the surface characteristics are usually unknown and varying, it is beneficial to extend the single-scale jet JEG towards a multi-scale representation JMSEG. For this jet the single-scale jets JEG_{ω_i}, obtained by filtering with Gabor filters of scales $\omega_1 \ldots \omega_M$, are simply concatenated,

$$\mathrm{JMSEG}(p) = [\mathrm{JEG}_{\omega_1}, \ldots, \mathrm{JEG}_{\omega_M}] \tag{4}$$

Self-Quotient Image (SQI): SQI was introduced by Wang et al. [17] as a method to separate the albedo information $R(x, y)$ from images. Similar to other works in this area, the idea is - based on the Lambertian assumption - that the illumination effects mainly appear in the low-frequency components of the image and that they can therefore be eliminated by dividing the image by a smoothed version of it. The method is intentionally designed for textured objects, but showed superior performance in a study of illumination invariance for face recognition [11] on the nearly textureless face parts cheek, chin and nose. Another motivation for including SQI in our evaluation is to assess the performance of the vast amount of methods dedicated to textured objects by evaluating one representative method.

For our experiments, we use the implementation of SQI provided by the INFace[1] toolbox and take the SSD of the SQIs as distance measure.

Gray Value (GV): In order to have a baseline performance, we also report results for simple image differencing, i.e., measuring the SSD between the original gray values of the two images.

3 Experiments

Experiments are conducted on synthetic image datasets built from 3D historical coin models as well as on real datasets of textureless and textured objects. We use synthetic images because this way the parameters of image formation can be freely changed to produce images with different illumination conditions and material properties with or without texture. In this manner, we are able to directly compare the performance of the features under different conditions without introducing a bias due to different objects used between datasets. The real dataset is used to validate our results for various real-world material properties and illumination conditions.

Synthetic Datasets. The synthetic datasets consist of images of 14 coin models which were rendered using the open-source graphics software *Blender*[2]. For each model, twelve sets of 500×500 images with 65 illumination directions were rendered where each set represents one out of four material BRDFs and one out of three texture density levels. Material BRDFs are intended to represent different levels of specularity starting from a Lambertian material with zero specularity up to specular intensity values of 0.25, 0.50 and 1.00. Three texture density levels were chosen to show the correlation of the features' performances to the amount of texture on the objects. The first level shows no texture and thus represents the set of textureless objects. For the remaining two levels we used synthetically generated textures. For each model and dataset, 65 images with varying illumination directions were rendered. The camera image plane is placed parallel to the coin and light source positions are defined by their azimuth angle ϕ and elevation angle λ. We used eight levels of λ with eight levels of ϕ each to produce 64 images. The 65th image is rendered with the light placed at the camera position (i.e. $\lambda = 90°$). Fig. 1(a) shows images of one model rendered with the same illumination parameters for the twelve synthetic datasets. Please note that the coin models have, on a local level, smooth isotropic as well as non-isotropic surface parts and thus cover the wide range of surface characteristics desired for our purpose. The dataset is available for download[3].

Amsterdam Library of Object Images. The Amsterdam Library of Object Images (ALOI) [10] is an image database of 1000 objects that were photographed from three viewpoints and with eight illumination configurations each.

[1] `http://luks.fe.uni-lj.si/sl/osebje/vitomir/face_tools/INFace/`
[2] `http://www.blender.org/`
[3] `http://www.caa.tuwien.ac.at/cvl/people/zamba/sidire/`

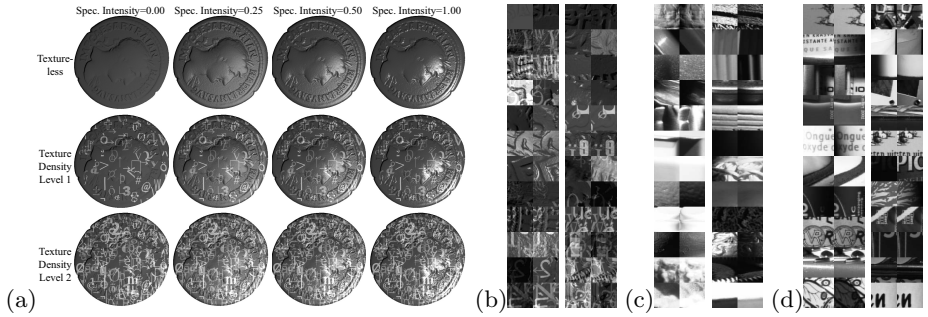

Fig. 1. (a) Coin model rendered with different material properties and texture densities; Patch pairs of size 64×64 from (b) the synthetic datasets, (c) ALOI textureless dataset and (d) ALOI textured dataset

The database contains a wide variety of textureless objects (e.g., a nut, a sponge, white cotton, a metal elephant, a plastic cup...) as well as textured objects (e.g., labeled boxes, an alarm clock, a calendar, a cream tube, a shoe ...). Therefore, the ALOI images provide a realistic and challenging database due to the high variation of material BRDF and surface smoothness among the objects.

3.1 Evaluation Scheme

The performance evaluation scheme is inspired by [4]: as a "good" feature will minimize the distance between image patches showing the same object part and maximize the distance between image patches showing different object parts, we measure these two groups of distances for a given feature and set of true and false image patch pairs. True patch pairs show the same object patch but with different illumination conditions, whereas false patch pairs show different object patches. False positive and true positive rates are sampled from these two groups of distances by means of a varying threshold and used to build a ROC curve of which the area under curve (AUC) is computed as performance measure. For patch pair generation we randomly extracted the same amount of true patch pairs and false patch pairs from the images of a dataset. We use a patch size of 16×16 pixels, but in general the patch size has no significant impact on the results, as has been observed in initial tests. Figure 1(b)-(d) show examples of true patch pairs from the synthetic datasets, the ALOI textureless dataset and the ALOI textured dataset, respectively (64×64 patches for better illustration). To generate patch pairs from the ALOI datasets, we manually identified 80 textureless objects and 80 textured objects in the dataset and randomly picked non-overlapping true and false patch pairs (12000 from the textureless objects and 18000 from the textured objects, as the textureless objects in the dataset are smaller on average).

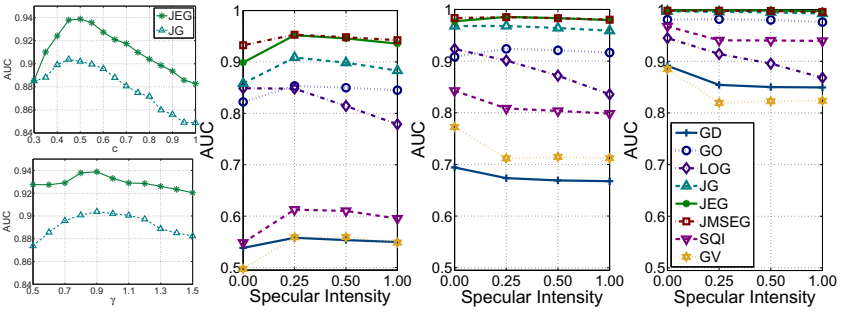

Fig. 2. (a) Recognition performance dependent on parameters c and γ; Recognition performance for different levels of material specularity on (b) textureless objects, (c) objects with texture density level 1 and (d) objects with texture density level 2

3.2 Results and Discussion

Parameter Selection. As our main purpose is the study of the features' behavior on textureless objects with varying material properties, tests for parameter selection were conducted on a mixed set consisting of 50000 true and false patch pairs extracted from the four synthetic datasets of textureless objects. Parameter selection was then achieved by an exhaustive search over the parameter space.

For GD and GO, we tested if a presmoothing of the patches or a bigger Sobel filter than the standard 3×3 one are beneficial in terms of recognition performance, but no improvement could be detected. For LOG we used an exhaustive search to find the optimal standard deviation of the Gaussian. For SQI, no exhaustive parameter selection was conducted as this method is intended for textured objects and initial tests with several parameter settings were not successful in substantially improving the generally bad performance of SQI. Therefore, we used the standard settings defined in the INFace Toolbox. For the features GJ and GEJ, the parameters defining the shape of the Gabor filters (c and γ) as well as the number N of orientations are of interest. Figure 2(a) shows the maximum AUC achieved over the parameter space for various fixed values of c and γ. We can derive from these results that the best performance is achieved when c is set in a range of $0.45 - 0.50$, i.e. the filters have a shape close to second derivatives of Gaussians. The optimal value for the aspect ratio of the filters defined by the parameter γ is around 0.9. Our experiments also revealed that the number of orientations does only have a minor influence on the overall performance for $N > 4$. Based on our results, for the further experiments we used parameter values of $\gamma = 0.9$ and $N = 6$ for JG, JEG and JMSEG, as well as $c = 0.50$ for JEG and JMSEG and $c = 0.45$ for JG. We also identified optimal filter sizes $\omega_1 \dots \omega_M$ for JMSEG as $\omega_j = \omega_1 2^{(j-1)/2}$ with $\omega_1 = 1$ and $M = 8$.

Recognition Performance Depending on Object Specularity and Texturedness. To evaluate the recognition performance of the features for the

twelve synthetic datasets, we randomly extracted 50000 true and false patch pairs for each dataset. Patch pairs contained in the mixed set for parameter selection are not contained in the four textureless datasets used for this evaluation. The results are plotted in Figure 2(b)-(d). It can it can be clearly seen that on textureless objects the representations based on even Gabor responses (JEG and JMSEG) perform best. The multi-scale representation of JMSEG is beneficial especially on Lambertian surfaces where it shows a significant improvement of recognition performance over JEG (AUC of 0.933 against 0.899). Complex Gabor filter responses (JG) are better than the other remaining features but it can be concluded from the worse performance compared to JEG that the phase invariance of the complex filter decreases its recognition power. For image gradients, there is a large discrepancy between the use of gradient orientations (GO) and gradient directions (GD). GD is much less stable than GO as it is highly vulnerable to edge polarity changes induced by opposite lighting directions between true patch pairs. SQI is only slightly better and performs substantially worse than the top-performing features, as this representation is designed for textured objects and is thus highly affected by changes of the shading patterns on textureless objects. Therefore, the method achieves its best results on Lambertian objects with a high texture density (Figure 2(d)). Another conclusion from the results on textureless objects is that more specularity of the objects' material increases the performance. Although a specular BRDF causes more appearance variations from light source variations than a Lambertian BRDF, surface characteristics are also more accentuated by a specular surface, which in turn supports its recognition. The only exception of this effect is LOG which has been especially proposed for smooth, Lambertian objects [15].

The results on textured objects shown in Figure 2(c)-(d) demonstrate that texture increases the recognition performance of all features and in general that their performance is correlated to the degree of texture variation, since albedo discontinuities are less affected by lighting variations than discontinuities of object depth. However, the representations based on Gabor filters are the best performing features for all material types, regardless of the texture density of the objects.

Influence of the Amount of Light Source Change. An interesting question in the context of our evaluation is how the amount of light source difference between the images to be compared has an influence on the discriminative power of the features. Therefore, we took this issue into account for the ROC curve generation by subselecting patch pairs from the textureless objects with a given difference of light source azimuth or elevation. Hence, only true patch pairs with a specified azimuth difference and no elevation difference, and vice versa, are considered. The results of these tests are shown in Figure 3(a)-(b). The plotted curves demonstrate that for smaller light source changes the performances of the features are close together whereas for stronger changes there is also a higher difference in performance. GD is a competitive feature for small light source changes of $45°$ azimuth and $10° - 20°$ elevation, but its performance decreases stronger than that of other features for larger light source changes. GD, SQI and

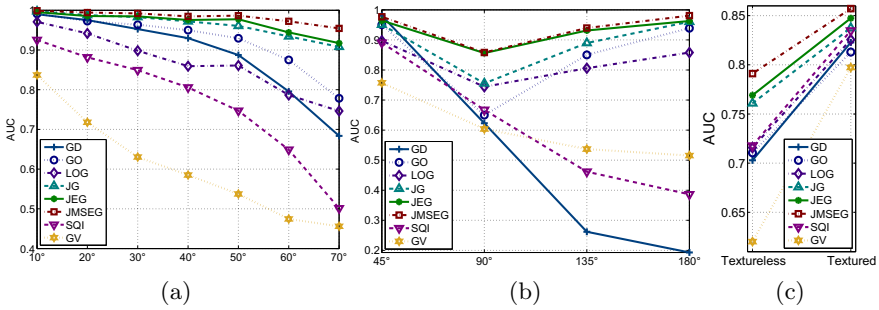

Fig. 3. Recognition performance in relation to (a) the difference of light elevation angle λ and (b) the difference of light azimuth angle ϕ; (c) Performance on real ALOI datasets

GV are especially vulnerable to changes of the light azimuth, as they are not invariant to edge polarities which tend to change on depth discontinuities for opposite lighting directions. An important aspect of these experiments is that the Gabor-based features show the top performance for all levels of light source changes, but their dominance is more pronounced for higher levels of change.

Recognition Performance on Real Datasets. As can be seen in Fig. 3(c), the results on the real datasets widely reflect the findings of the experiments on the synthetic datasets. JMSEG is again the best performing feature for textureless and textured objects, followed by JEG and JG. The generally lower performance on the real datasets is explained by image noise on the images as well as the acquisition setup used. There are more underexposed (i.e. completely black) and overexposed (i.e. completely white) objects parts which evidently hinders recognition. Nonetheless, the results show that the insights gained from our experiments on synthetic datasets can be transferred to the real world.

4 Conclusions

In this paper we addressed the problem of comparing images of textureless 3D objects in unconstrained conditions. Although in general invariance to illumination conditions is a central computer vision problem, we identified the subarea of textureless objects as an under-researched topic. Therefore, we evaluated several well-known pixelwise low-level features with respect to their recognition performance for textureless objects. Unlike previous studies [5,15], we separately evaluated textureless objects and demonstrated that features claimed to be insensitive to illumination conditions, like gradient direction or the self-quotient image, perform substantially worse on textureless surfaces than on textured surfaces. Our findings are based on a comprehensive evaluation on synthetic datasets with varying degrees of specularity and texturedness as well as real images of textureless and textured objects.

Our experiments revealed that jets of even Gabor responses are the features of choice for capturing surface characteristics in an illumination-insensitive way,

not only for textureless but also for textured objects under strong illumination changes. We also demonstrated that for improved performance one can extend the single-scale representation towards multiple scales by concatenating the single-scale jets. We think that this representation offers a powerful basis for more sophisticated methods that tackle computer vision problems involving textureless objects or heavily changing illumination conditions. For future work, we plan to intensively investigate their usage for higher-level features which will allow for the improved recognition, registration or reconstruction of objects in such scenarios.

References

1. Adini, Y., Moses, Y., Ullman, S.: Face recognition: The problem of compensating for changes in illumination direction. PAMI 19(7), 721–732 (1997)
2. Barrow, H., Tenenbaum, J.: Recovering intrinsic scene characteristics from images. Computer Vision Systems, 3–26 (1978)
3. Basri, R., Jacobs, D.: Lambertian reflectance and linear subspaces. PAMI 25(2), 218–233 (2003)
4. Brown, M., Hua, G., Winder, S.: Discriminative learning of local image descriptors. PAMI 33(1), 43–57 (2011)
5. Chen, H., Belhumeur, P., Jacobs, D.: In search of illumination invariants. In: CVPR, pp. 254–261 (2000)
6. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR, pp. 886–893
7. Freeman, W., Adelson, E.: The design and use of steerable filters. PAMI 13(9), 891–906 (1991)
8. Gabor, D.: Theory of communication. Journal of the Institute of Electrical Engineers 93, 429–457 (1946)
9. Georghiades, A., Belhumeur, P., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. PAMI 23(6), 643–660 (2001)
10. Geusebroek, J., Burghouts, G., Smeulders, A.: The amsterdam library of object images. IJCV 61(1), 103–112 (2005)
11. Gopalan, R., Jacobs, D.: Comparing and combining lighting insensitive approaches for face recognition. CVIU 114(1), 135–145 (2010)
12. Kamarainen, J., Kyrki, V., Kalviainen, H.: Invariance properties of gabor filter-based features-overview and applications. TIP 15(5), 1088–1099 (2006)
13. Lowe, D.: Distinctive image features from scale-invariant keypoints. IJCV 60(2), 91–110 (2004)
14. Moreels, P., Perona, P.: Evaluation of features detectors and descriptors based on 3d objects. IJCV 73(3), 263–284 (2007)
15. Osadchy, M., Jacobs, D., Lindenbaum, M.: Surface dependent representations for illumination insensitive image comparison. PAMI 29(1), 98–111 (2007)
16. Van De Sande, K., Gevers, T., Snoek, C.: Evaluating color descriptors for object and scene recognition. PAMI 32(9), 1582–1596 (2010)
17. Wang, H., Li, S., Wang, Y.: Face recognition under varying lighting conditions using self quotient image. In: FG, pp. 819–824 (2004)