Robust point matching in HDRI through estimation of illumination distribution

Yan Cui   Alain Pagani   Didier Stricker

DFKI, Augmented Vision
Kaiserslautern University, Germany

Abstract. High Dynamic Range Images provide a more detailed information and their use in Computer Vision tasks is therefore desirable. However, the illumination distribution over the image often makes this kind of images difficult to use with common vision algorithms. In particular, the highlights and shadow parts in a HDR image are difficult to analyze in a standard way. In this paper, we propose a method to solve this problem by applying a preliminary step where we precisely compute the illumination distribution in the image. Having access to the illumination distribution allows us to subtract the highlights and shadows from the original image, yielding a material color image. This material color image can be used as input for standard computer vision algorithms, like the SIFT point matching algorithm and its variants.

1   Introduction

While High Dynamic Range Images (HDRI) representing the real word's range of luminance are commonly used in the Computer Graphics community, their use in machine vision tasks (e.g. Registration and Identification) is not widespread in the Computer Vision community. HDRI can measure a high radiance and illumination range for the real world scenes, thus providing more information than the low dynamic range images (LDRI). Many applications, such as image-based lighting [6], and BRDF measurement [14] require access to the whole dynamic range of a scene. In this paper, we present a method to use HDR images for computer vision tasks by estimating the illumination distribution first and applying a suitable tone-mapping method for computer analysis. We apply this concept to the matching problem, yielding a SIFT [16] method for the HDRI.

Illumination distribution estimation is an important task for the computer vision. The appearance of objects depends greatly on illumination conditions. Since substantial image variation can result from shading, shadows and highlights, there has been much research on dealing with such lighting effects for a LDRI [3] [15] [13], but not much for HDRI [23]. Because of the significant effect of lighting, it is often helpful to know the lighting conditions of a scene so that an image can be more accurately analyzed. Recovery of illumination conditions is also important for computer graphics applications, such as inserting correctly shaded virtual objects into augmented reality systems [22] and lighting reproduction for compositing actors into video footage [7]. While these graphics methods

introduce special devices into a scene to capture the lighting distribution, estimation of illumination in image has proven to be a challenge. In this paper, we do not only estimate light source position (like e.g. [23]), but we provide the illumination distribution as a Gaussian Mixture Model (GMM) over the image for each different exposure layer in the HDR image.

The development of techniques for HDRI capture and synthesis have made tone-mapping an important problem in computer graphics [9]. The fundamental problem is how to map the large range of intensities found in an HDRI into the limited range generated by a conventional display device. There are three main taxonomies of tone-mapping operators. A primary distinction is whether an operator is global or local. Global operators apply a single mapping function to all pixels of the image, whereas local operators modify the mapping depending on the characteristics of different parts of the image. A second important distinction is between empirical and perceptually based operators. A third distinction is between static and dynamic operators. In this paper we suggest to use a new tone-mapping method, more suited to computer vision tasks, and to get a "material color" of the scene or the object without the illumination interference.In all feature extraction methods, the invariance with respect to imaging conditions represents the biggest challenge. More specifically, the local extracted features should be invariant with respect to geometrical variations, such as translation, rotation, scaling, and affine transformations. Furthermore, these features should be invariant with respect to photometric variations such as illumination direction, intensity, colors, and highlights. SIFT [16] [17] has been proven to be the most robust among the local invariant feature descriptors with respect to different geometrical changes [20]. However, due to the color constancy problem, a lot of geometrical invariant approaches avoid dealing with illumination problem. Therefore, illumination invariance is a crucial problem which has to be solved for local features. While some researchers already focused on the color constancy problem [2] [19], some attempts to make use of the color information inside the SIFT descriptors have been proposed [5] [4] [10] [1]. In this paper, we solve the illumination invariance problem for the HDR images, using result of our illumination distribution estimation.

The paper provides three main contributions to the HDR image processing research. First, we show that it is possible to estimate the illumination distributions in each exposure layer of HDRI with a Gaussian Mixture Model. Second, we propose a new tone-mapping algorithm which is more suitable for the computer analyzing through material color recovery. Third, as an application, we show that the SIFT algorithm using the tone-mapped images performs better in terms of robustness and number of matches.

The remainder of the paper is organized as follows: We first present illumination distribution with GMM in Sect. 2. In Sect. 3, we explain how to estimate the GMM parameters in HDRI. We show how to recover the shadow and highlight parts position in the image from the illumination distribution result in Sect. 4. Finally, the SIFT method for HDRI and the results is presented in Sect. 5. We conclude in Sect. 6 with directions for future work.

2   Illumination distribution with GMM

We can assume that the illumination distribution in the 2D image is a Gaussian mixture model (GMM) for several light sources. For a single light source, we assume that the illumination distribution is a Gaussian model as Eq. (1)

$$e(\boldsymbol{x}) = p(\boldsymbol{x}|\,\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{\mu})^T(\boldsymbol{x}-\boldsymbol{\mu})}{2\sigma^2}\right) \tag{1}$$

In the Eq. (1), $\boldsymbol{x}$ is the 2d image position, $\boldsymbol{\mu}$ is the light source position, σ stands for the light intensity and the distribution property, and we assume that σ is same for the two direction, but different for each exposure layer in the multi-exposure sequence of the HDRI. There are 11 exposure layers created from one light source HDRI, as shown in Fig. 1a, the exposure times varied by powers of two between f-stops from $\frac{1}{32}$ to 32. Fig. 1b shows the Gaussian illumination distribution results for each exposure layer with the method in Sect. 3, the light source position $\boldsymbol{\mu}$ is not changed, the variance σ increases from layer to layer.

Then we can assume that the GMM model for more than one light sources, as the Eq. (2)

$$e(\boldsymbol{x}) = \sum_{k=1}^{K} \rho_k * p(\boldsymbol{x}|\,\theta_k) \quad with \quad \boldsymbol{x}|\,\theta_k \propto N(\boldsymbol{\mu}_k, \sigma_k) \tag{2}$$

In the Eq. (2), ρ_k is the mixture weight for the light source k, we assume that ρ_k are same for the different exposure layer to the same light source. Fig. 1c shows one exposure layer created from a two light sources HDRI, Fig. 1d expresses the Gaussian distribution result as the method in Sect. 3.

Our hypothesis is that illumination can be estimated as a GMM for different exposure layers.

3   Illumination distribution in HDRI

In this section we show how to estimate the GMM parameters for the illumination distribution in each exposure layer. The input of our algorithm is a HDRI. We first create a multi-exposure durations sequence with normal global tone-mapping method. Equivalently, the input can be a number of digitized photographs taken from the same point with different known exposure durations t_j. Using diffuse body reflection Lambertian model [11]:

$$E_i = e(i)R_\infty\,(\lambda, i) \tag{3}$$

In the Eq. (3), where i denotes the position at the imaging plane and λ the wavelength. Further, $E(\lambda, i)$ denotes the illumination spectrum, the material reflectivity is denoted by $R_\infty\,(\lambda, i)$, for this part, it's the property of the material, we call it "material color", which is used in the feature extraction in Sect. 5.

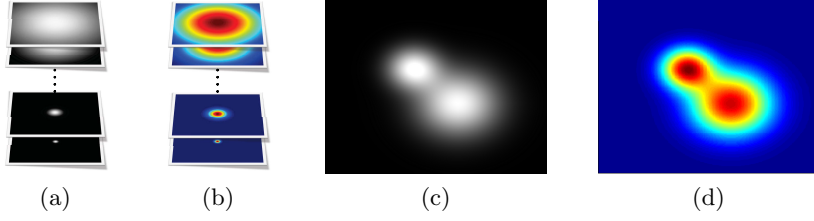For different exposure time layer j, as [8], we can get:

Fig. 1: (a) A multi-exposure sequence of HDRI with one light source. (b) Gaussian distribution results for (a); (c) One HDRI exposure layer with two light sources. (d) Gaussian distribution results for (c).

$$M_{ij} = e(i)R_\infty\left(\lambda, i\right)t_j \tag{4}$$

As presented in Sect. 2, we can estimate the $e(i)$ part as GMM for the illumination distribution. Taking the GMM $e(i)$ to the Eq. (4), the final illumination distribution function for each pixel in each layer:

$$M_{ij} = \left(\sum_{k=1}^{K} \rho_k * p(i|\theta_{k,j})\right) R_\infty\left(\lambda, i, j\right)t_j \tag{5}$$

Finally, the "material color" $R_\infty\left(\lambda, i, j\right)$ part for each pixel can be expressed:

$$R_\infty\left(\lambda, i, j\right) = M_{ij} / \left(\sum_{k=1}^{K} \rho_k * p(i|\theta_{k,j})t_j\right) \tag{6}$$

For each exposure layer j, we can assume that the "material color" $R_\infty\left(\lambda, i, j\right)$ part is the same. We define the energy function:

$$E\left(\rho_k, \theta_{k,j}\right) = \sum_{j=1}^{L-1}\sum_{i=1}^{N}\left(\frac{M_{ij+1}}{\sum_{k=1}^{K}\rho_k p(i|\theta_{k,j+1})t_{j+1}} - \frac{M_{ij}}{\sum_{k=1}^{K}\rho_k p(i|\theta_{k,j})t_j}\right)^2 \tag{7}$$

In Eq. (7), L stands for the number of different exposure duration layers, N is the number of pixels, and K is the number of the light source for each layer. We assume that the light intensity σ in θ are different, but the light position μ in θ and the light intensity weight ρ are same for each layer. We use iterative Expectation Maximization (EM) procedure to find a solution to minimize the energy function Eq. (7). Because $\sum_{k=1}^{K}\rho_k p(i|\theta_{k,j})t_j \neq 0$ for each pixel in every layer, then energy function turns to:

$$E\left(\rho_k, \theta_{k,j}\right) = \sum_{k=1}^{K}\sum_{j=1}^{L-1}\sum_{i=1}^{N}\rho_k\left(p(i|\theta_{k,j})t_j M_{ij+1} - p(i|\theta_{k,j+1})t_{j+1}M_{ij}\right)^2 \tag{8}$$

With the EM algorithm, we can get the Q function:

$$Q\left(\rho,\theta\right) = \sum_{k=1}^{K} \sum_{j=1}^{L-1} \sum_{i=1}^{N} \rho_{k}^{old} \left(\ln(p(i|\theta_{k,j})t_{j}M_{ij+1}) - \ln(p(i|\theta_{k,j+1})t_{j+1}M_{ij})\right) \quad (9)$$

$$\rho_{k}^{old} = \frac{\sum_{j=1}^{L-1} \sum_{i=1}^{N} \left(\ln(p(i|\theta_{k,j})t_{j}M_{ij+1}) - \ln(p(i|\theta_{k,j+1})t_{j+1}M_{ij})\right)}{\sum_{k=1}^{K} \sum_{j=1}^{L-1} \sum_{i=1}^{N} \left(\ln(p(i|\theta_{k,j})t_{j}M_{ij+1}) - \ln(p(i|\theta_{k,j+1})t_{j+1}M_{ij})\right)} \quad (10)$$

We set the initial value $\rho_{k}^{old} = 1$. During the minimization-step, we can estimate $\sigma_{k,j}$ to maximum Q function. Then during expectation-step, we estimate the new ρ_{k} for each alternative light source position. If the weight $\rho_{k} < T_{light}$ (T_{light} is a threshold defined by user), we consider this pixel is not the light source position, and assign 0 to this weight ρ_{k} directly. For the experiments, we set $T_{light} = 4.0$.

The above EM procedure converges to a local minimum of the Eq. (7). Please note that the variances $\sigma_{k,j}$ are continuously recomputed, they're increasing from low exposure layer to high exposure layer, which is similar to an annealing procedure in which support of the Gaussians is reduced when assignment 0 to this weight ρ_{k}.

The first result are shown in Fig. 1b for one light source and Fig. 1d for two light sources. For these two experiments, the lights and the scene are controled strictly, the exposure times varied by powers of two between f-stops from $\frac{1}{32}$ to 32. K is the alternative light sources, considering as the whole image size $1...N$. We can see the illumination changed for one light source in Fig. 1b, and the illumination distribution for two light source in Fig. 1d.

In order to make the EM convergence fast, the initial light area can be calculated firstly, then the alternative light source number K in the energy function becomes smaller than the whole image size. For the initial light area detection for HDRI, there are known methods to solve this problem [23] [12]. However, for our case, we need not to estimate the light source accurately. We can get the initial light source from the low exposure time layer with a threshold. In the low exposure time layer, the light densities are low, and if the pixel value is bigger than a threshold, we can consider it as the initial alternative light position. When the energy function is minimized with EM algorithm step by step, an accurate light source position will be determined by the weight parameters ρ_{k}.

We test our approach for natural environments and complex light conditions in two scenes: Church and Studio.

For the Church scene, a number of digitized photographs are taken from the same point with different known exposure durations, there are 16 photographs of a church taken at 1-stop increments from 1/1000 sec to 30 sec. Fig. 4 shows 5 samples from the sequence, the exposure times are 0.0146, 0.1172, 0.4688, 1.8750, 30.0 sec. The sun is directly behind the rightmost stained glass window, making it especially bright. The initial light source is given by exposure 0.1172 layer, as Fig. 2a. The final illumination distribution for these layers results are shown in Fig. 5. The algorithm can detect three main light sources, window on the top of
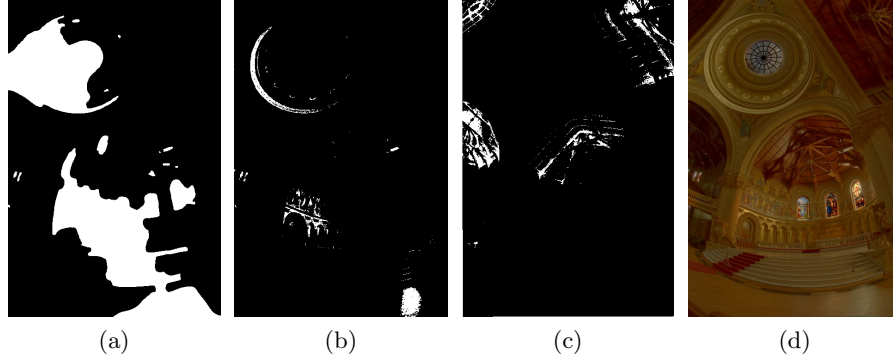
Fig. 2: (a)Initial light sources for EM procedure. (b) Highlight and (c) Shadow area for church of exposure time 1.8750 sec layer. (d)"Material color"

the church and three windows in the middle of the church. As the exposure time increase, the illumination distribution is changed, but the light source position is not changed. The light intensity is increasing layer to layer.

For the Studio scene, we can created 11 exposure layers by normal global tone-mapping method, the exposure time increments from $-11EV$ to $2EV$. As Fig. 6 shows 5 samples from the sequence, the exposure times are -10, -8, -1.5, 0, 1.5 EV. The sun is outside the glass window, making it especially bright. The initial light source is given by exposure $-10EV$ layer, as Fig. 3a. The final illumination distribution for these layers results are shown in Fig. 7. The algorithm can detect the light sources from outside of the window and the lamp inside the room.

From the experiments, our algorithm not only can detect the point light sources as the Fig. 1b and Fig. 1d, but also can detect the plane light source as the natural case, as Fig. 5 and Fig. 7. Once we have computed the illumination distribution information, we can compute highlight parts and shadow parts in the image, as described in the Sect. 4

4   Shadow and highlight parts in HDRI

In this section, the shadow, highlight parts and the "material color" images are calculated by the illumination distribution results in Sect. 3. In Eq. (4), $e(i)$ is the light sources distribution, in the real environment image, we can assume that there is another light source (ambient light) that distributes evenly in the 2D image. For our experiments, the ambient light A is constant, $A = 0.001$. We can then derive the "material color" for each layer:

$$R_\infty (\lambda, i, j) = M_{ij} / \left((e(i) + A) \, t_j \right) \tag{11}$$

For each layer, we define two thresholds T_{up} and T_{below}, If the "material color" $R_\infty (\lambda, i, j) > T_{up}$, consider this area as the highlight part for this layer,

Fig. 3: (a)Initial light sources for EM procedure. (b) Highlight and (c) Shadow area for Studio of exposure 0 EV layer. (d)"Material color"

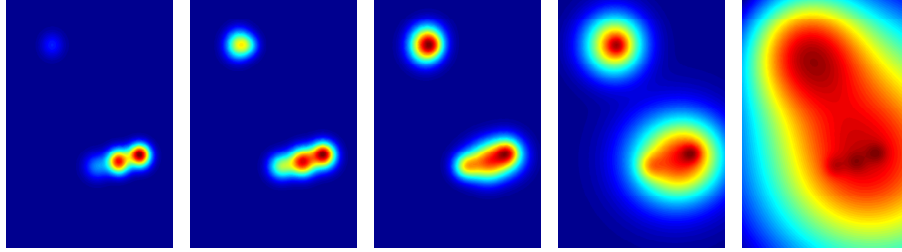Fig. 4: Church, exposure times sequence 0.0146, 0.1172, 0.4688, 1.8750, 30.0 sec.

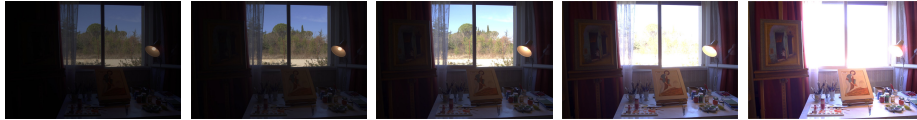Fig. 5: Church, illumination distribution results for the image layers above.

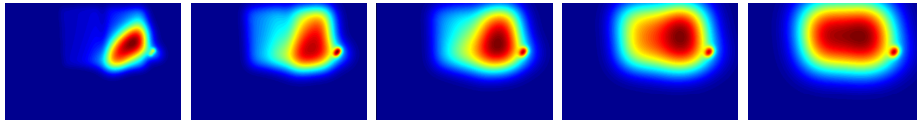Fig. 6: Studio, exposure times sequence −10, −8, −1.5, 0, 1.5 EV

Fig. 7: Studio, illumination distribution results for the image layers above.

the highlight part are not included the light source positions. Similarly, if the "material color" $R_\infty\left(\lambda, i, j\right) < T_{below}$, we consider this area as the shadow part for this layer. For the experiments, we set $T_{up} = 240$ and $T_{below} = 20$. The highlight parts and shadow parts for the church exposure time 1.8750 sec layer are shown in Fig. 2b 2c; The result for the studio exposure $0EV$ is shown in Fig. 3b 3c. Furthermore, the "material color" can be determined by $R_\infty\left(\lambda, i, j\right)$ in multi-exposure layers $j = 1...L$:

$$R_\infty\left(\lambda, i\right) = \left(\sum_{j=1}^{L} w(j)R_\infty\left(\lambda, i, j\right)\right) / \sum_{j=1}^{L} w(j) \tag{12}$$

Because the information in the middle exposure time layers are more reliable, the weight parameters $w(j)$ are assigned as 1-D Gaussian distribution, $w(j)$ are small for the lower and higher exposure time, and large for the middle exposure time in the sequence. We can consider this step as a tone-mapping procedure. It is worth noting that, different from the usual tone-mapping methods that try to produce a pleasant visual effect, our tone mapped results are more suitable for the computer analysis. We test the image feature extraction and corresponding problem with SIFT with our tone-mapping algorithm in Sect. 5. The final "material color" image results are shown in Fig. 2d of church image and Fig. 3d of studio image. There is no highlight area in the final image.

5   Point matching in HDR images and results

In this section we will calculate the corresponding points with SIFT method for two HDRI for a same scene and a same object. As Sect. 4, the "material color" image can be calculated from one HDRI, without highlight area. The invariant features can be detected. We use PC-SIFT method [5], which is motivated by perception-based color space, instead of using the gray value as input image, PC-SIFT approach builds SIFT descriptors in 3 channel color space, is more robust than normal SIFT with respect to color and photometrical variations.

The main stages using local invariant features are interest points detection, descriptor building and descriptor matching. Interest points should be selected so that they achieve the maximum possible repeatability under different photometric and geometric imaging conditions. As discussed in Sect. 3, our SIFT is based on "material color" image, which is illumination invariance. In the same time, the extrema in Laplacian pyramid, which is approximated by difference-of-Gaussian for the input image in different scales, has been proven to be the most robust interest points detector to geometrical changes [20] [4].

The experiment results are shown below. First, we test an object feature with two different views, for each view, there are 12 exposure duration from $\frac{1}{60}$ sec to 0.4 sec. In Fig. 8a shows the $\frac{1}{8}$ sec layer and the PC-SIFT result. The "material color" and the PC-SIFT results are shown in Fig. 8b, from which we can notice there is no highlight and shadow part on the object. Meanwhile the SIFT can find the corresponding points without the light interference.
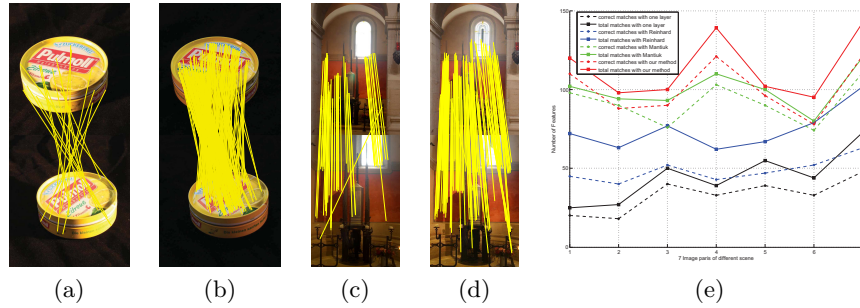
Fig. 8: (a) One exposure time 1/8 sec and (b) 0.8EV layer of 12 multi-exposure and PC-SIFT result, (b)(d) The "material color" and PC-SIFT result. (e)Total matches and correct matches of different tone-mapping methods.

Second, we test a scene with two different views HDRI, 14 exposure layers are created by normal tone-mapping method from −3EV to 9EV, In Fig. 8c, shows the 0.8EV layer and the PC-SIFT result. The "material color" and the PC-SIFT results are shown in Fig. 8d, there is not enough information to clear the shadow part, because the shadow parts exist in each exposure layer. For the SIFT result, our algorithm can detect the corresponding without the light source factor, finding the correct corresponding near the window part. Further, the shadow part can be extracted, as Fig. 2c shows, if we do not like the corresponding points in the shadow area, they can be cleaned away.

Finally, we compare our method (red line) to the other tone-mapping algorithm, global tone-mapping filter Reinhard [21] (blue line), local filter Mantiuk [18] (green line) and one exposure LDR image layer (black line) in Fig. 8e, the dotted lines show the correct matches, and the solid lines show the total matches for the 7 image paris of different scenes. As the result show, our method can detect more matches and find more correct matches.

6 Conclusions

In this paper, we presented a robust point matching approach for HDR images. Our method is based on a robust estimation of the illumination distribution in the 2D image using a Gaussian Mixture Model. The parameters of the GMM are recovered directly from the HDRI with EM-algorithm. With the estimated illumination distribution in the 2D image, we can compute the highlight parts, shadow parts and "material color" image, which is suitable for many computer vision tasks. We show that we can successfully apply this method to the point matching problem, using SIFT as underlying method. Our results show that a better matching is achieved in terms of robustness and number of matches.

References

1. Abdel-Hakim, A.E., Farag, A.A.: Csift: A sift descriptor with color invariant characteristics. In: CVPR, 2006 IEEE. vol. 2, pp. 1978–1983 (2006)
2. Brainard, D., Freeman, W.: Bayesian color constancy. the Journal of Optical Society of America 14, 1393–1411 (1997)
3. Brooks, M.J., Horn, B.K.P.: Shape from shading. chap. Shape and source from shading, pp. 53–68. MIT Press, Cambridge, MA, USA (1989)
4. Brown, M., Lowe, D.G.: Invariant features from interest point groups. In: In British Machine Vision Conference. pp. 656–665 (2002)
5. Cui, Y., Pagani, A., Stricker, D.: Sift in perception-based color space. In: IEEE 17th International Conference on Image Processing (ICIP). pp. 3909 – 3912 (2010)
6. Debevec, P.: Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. In: ACM SIGGRAPH 2008. pp. 32:1–32:10. New York, NY, USA (2008)
7. Debevec, P., Wenger, A., Tchou, C., Gardner, A., Waese, J., Hawkins, T.: A lighting reproduction approach to live-action compositing. ACM Trans. 21, 547–556 (2002)
8. Debevec, P.E., Malik, J.: Recovering high dynamic range radiance maps from photographs. pp. 369–378. SIGGRAPH '97, New York, NY, USA (1997)
9. Devlin, K., Chalmers, A., Wilkie, A., Purgathofer, W.: Star: Tone reproduction and physically based spectral rendering. In: dcwp (ed.) State of the Art Reports, Eurographics 2002. pp. 101–123. The Eurographics Association (September 2002)
10. Farag, A., Abdel-Hakim, A.E.: Detection, categorization and recognition of road signs for autonomous navigation. ACIVS2004 pp. 125–130 (2004)
11. Judd, D.B., Wyszecki, G.: Color in Business, Science, and Industry. New York
12. Krawczyk, G., Mantiuk, R., Myszkowski, K., Seidel, H.P.: Lightness perception inspired tone mapping. In: Proceedings of the 1st Symposium on Applied perception in graphics and visualization. pp. 172–172. ACM, New York, NY, USA (2004)
13. Lalonde, J.F., Efros, A.A., Narasimhan, S.G.: Estimating natural illumination from a single outdoor image. In: IEEE ICCV (2009)
14. Lensch, H.P.A., Kautz, J., Goesele, M., Heidrich, W., Seidel, H.P.: Image-based reconstruction of spatial appearance and geometric detail. ACM Trans. Graph. 22, 234–257 (April 2003)
15. Li, Y., Lin, S., Lu, H., yeung Shum, H.: Multiple-cue illumination estimation in textured scenes. In: IEEE Proc. 9th ICCV. pp. 1366–1373 (2003)
16. Lowe, D.G.: Object recognition from local scale-invariant features. In: Computer Vision, 1999. vol. 2, pp. 1150–1157 vol.2 (1999)
17. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60(2), 91–110 (2004)
18. Mantiuk, R., Myszkowski, K., Seidel, H.P.: A perceptual framework for contrast processing of high dynamic range images (2005)
19. M.D'Zmura, Lennie, P.: Mechanisms of color constancy. the Journal of Optical Society of America 3, 1662–1672 (1986)
20. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. PAMI, IEEE Transactions on 27(10), 1615–1630 (Oct 2005)
21. Reinhard, E., Stark, M., Shirley, P., Ferwerda, J.: Photographic tone reproduction for digital images. In: PROC. OF SIGGRAPH02. pp. 267–276. ACM Press (2002)
22. Sato, I., Sato, Y., katsushi Ikeuchi: Acquiring a radiance distribution to superimpose virtual objects onto a real scene (1999)
23. Yoo, J.D., Cho, J.H., Kim, H.M., Park, K.S., Lee, S.J., Lee, K.H.: Light source estimation using segmented hdr images. SIGGRAPH '07, ACM, NY, USA (2007)