

Color Target Localization under Varying Illumination Conditions

Simone Bianco and Claudio Cusano

DISCo (Dipartimento di Informatica, Sistemistica e Comunicazione),
Università degli Studi di Milano-Bicocca, Viale Sarca 336, 20126 Milano, Italy
{bianco,cusano}@disco.unimib.it

Abstract. In this work we have investigated the use of color descriptors to automatically locate the color target in the scene. Three different local descriptors have been tested. These descriptors are then used to return multiple localization hypotheses and a geometrical and appearance validation are introduced to select the most feasible pose. The experimental results on a public dataset of RAW images containing the Macbeth ColorChecker CC target and acquired in uncontrolled environments, showed that all the descriptors considered benefited from the hypothesis validation introduced.

1 Introduction

Color targets are widely used in imaging applications. There are two approaches in the field of input device colorimetric characterization: the spectral sensitivity based approach and the color target based approach. In the former approach, the input device spectral sensitivity needs to be measured using expensive specialized apparatus [1]. In the latter approach only a known target is required, making it more practical [2,3]. In the field of multispectral imaging the color targets are also crucial [4]: in fact, for the system characterization a color target with known reflectances is needed. Common practice in the computational color constancy field is to place a known color target in the scene to measure the ground truth illuminant and to assess the performance of the computational color constancy algorithms [5]. Color targets are also used to measure the physical properties of the devices, such as the dynamic range, the noise levels and the optical resolution.

Different color targets are commercially available, each one specifically designed for a different class of devices and for the property to assess (see Figure 1 for some examples). Moreover, custom targets can be used for specific applications: for example Smoyer et al. [6] used a custom made Gamblin oil paint target to assess the reproduction quality of museum digital camera systems; Marguier et al. [7] used a custom made color target to assess human skin color from uncalibrated images.

The color targets placed in the scene are usually manually localized to extract the required information. When dealing with large image datasets, as for example those used for computational color constancy, the target localization can

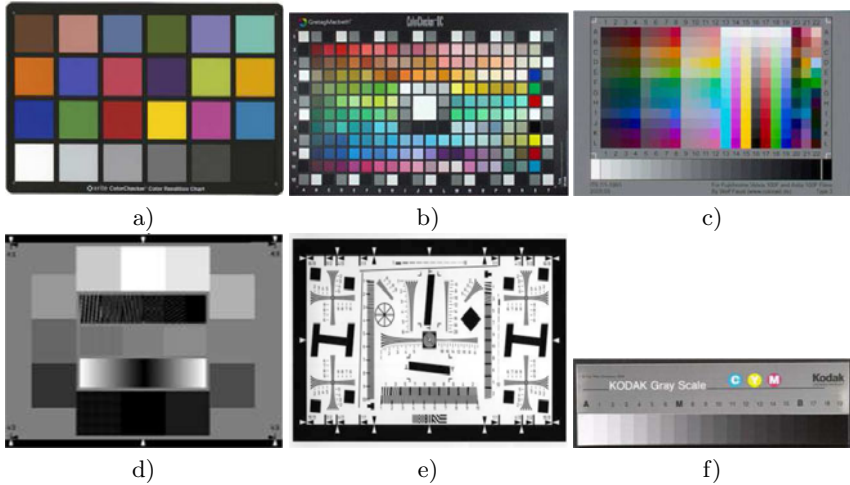


Fig. 1. Examples of standard targets used in imaging applications: a) Macbeth ColorChecker CC; b) Macbeth ColorChecker DC; c) Agfa IT8; d) ISO 15739 noise test chart; e) ISO 12233 resolution test chart; f) Kodak gray scale

be lengthy and tedious. To overcome this problem, we designed a method, based on local color descriptors, to automatically locate the color target in the scene. Three different local descriptors have been tested for the localization task of the Macbeth ColorChecker CC (MCC) in a dataset of 568 RAW images acquired in uncontrolled environments [5]. These descriptors are then used to return multiple localization hypotheses which are then validated on the basis of geometrical and appearance criteria which select the most feasible pose. Appearance validation has been specifically designed to be robust with respect to variations in illumination conditions.

The paper is organized as follows. In Section 2 the general approaches for object localization are described. In Section 3 the proposed approach is described. In Section 4 the experimental results are reported. Finally, in Section 5, conclusions are drawn.

2 Object Localization Methods

Object detection consists in finding instances of a given class of objects in digital images and videos. Object localization is a simplified scenario where it is assumed that exactly one instance is present in the scene. In many cases the class of objects of interest is simply defined by a single image showing a reference instance of the objects to be found. Template matching techniques represent a simple approach to object detection and localization. While very effective for certain environments where object pose and illumination are carefully controlled, template matching

becomes infeasible when significant variations in scale, illumination, and 3D pose are allowed. This approach is even more problematic when the object is partially occluded, or placed in a cluttered scene. An alternative approach is to consider only particularly distinctive parts of the image which can be easily matched. This approach usually consists of three major steps [8]: i) a set of distinctive keypoints is selected, ii) the patches surrounding each detected keypoint are described, iii) the set of descriptors extracted from the target and from the template image are compared and matched. Corner detectors (such as the Harris corner detector [9]) are a common choice for keypoint extraction. They identify those locations corresponding to peaks in local image variation which can be reliably and accurately identified in different images. A common problem of early corner detectors is that they examine an image at a single scale: different keypoints are detected when the scale changes significantly. To overcome this problem a multi-scale detector is needed. For instance, Lowe developed the Scale Invariant Features Transform (SIFT) [10], which analyzes the differences of Gaussians at multiple scales. The resulting extrema in the scale-space are selected and labeled with the corresponding scale, and with an orientation computed on the basis of local gradient directions. For each keypoint a neighborhood is selected on the basis of the location, scale and orientation and is described by a feature vector. In order to allow a robust matching, these descriptors should be invariant with respect to changes in illumination and viewpoint. The local descriptors considered in this work are illustrated in the following section. In order to perform object detection, the keypoints and the descriptors found on both the target and the template image must be matched. The approach considered here is similar to the one originally proposed by Lowe [10]. For each keypoint of the template, the best match in the target image is computed by comparing all the extracted descriptors. In case of ambiguity (high similarity for multiple target descriptors, or low similarity for all the target descriptors) the pair is ignored. An affine transform is determined to geometrically match the template and the target keypoints of each pair of matching descriptors. To do so, the Hough transform is used to identify clusters of matching pairs, and each cluster is then verified by a least-squares fitting of the best affine projection parameters relating the template image to the target image. In our approach, the largest clusters are further validated by considering geometric and appearance criteria. The most suitable hypothesis (i.e. the cluster which matches best those criteria) is finally selected.

3 The Proposed Approach

The proposed approach involves the following steps:

1. Detection of scale-space extrema using a difference-of-Gaussian function to identify potential interest points that are invariant to scale and orientation.
2. A detailed model is fit to determine location and scale at each candidate location. Only stable keypoints are retained.

3. On the basis of the local image gradient directions one or more orientations are assigned to each keypoint location.
4. Taking into account the keypoint scale and orientation, each of the descriptors considered is used to describe the region around each keypoint.
5. Once all the keypoints of the color target and of the image in which we want to locate it are described, they are matched to find corresponding keypoints.
6. A Hough transform is used to identify the largest clusters of keypoints belonging to a single object.
7. For each of the identified clusters a verification step is computed through least-squares solution to find consistent pose parameters.
8. Pose parameters are verified to ensure that they correspond to admissible position and orientation.
9. Color descriptors, extracted from the template and the region identified by the pose parameters, are compared.
10. The region which maximizes the similarity with the template is finally selected.

The whole procedure, with the exception of the last three steps, is general purpose and could be applied to various object detection/localization problems. The main difference with the standard approach is that in step (6) the Hough transform is not used to identify a single cluster of keypoints, but it is used to identify multiple clusters. This means that we generate multiple hypotheses about the target pose and we need an hypothesis validation step to select the best hypothesis. Hypothesis validation is performed by the last three steps, which have been specifically designed for the target localization problem by taking into account the following considerations: i) the color target must be placed in such a way that it is entirely contained in the image, and must be oriented in such a way that it is clearly visible; ii) the color target can be easily characterized in terms of its color distribution; iii) a color correction method should be applied to make the color descriptor robust with respect to changes in illumination.

3.1 Local Descriptors Considered

The color descriptors considered belong to the class of gradient-based descriptors. The first one is the Scale Invariant Feature Transform (SIFT): The SIFT descriptor has been proposed by Lowe [10] and describes the local shape of a region using edge orientation histograms. The region is described using a 4×4 array of histograms with 8 orientation bins in each, giving a descriptor with length $4 \times 4 \times 8 = 128$. The second and third descriptors considered are extensions of the SIFT: they are the Opponent-SIFT and the rg-SIFT. The Opponent-SIFT has been included as it is suggested in [11] for its ability to cope with unknown data. The rg-SIFT has been included as it performed best in a previous target localization task [12].

The Opponent-SIFT descriptor computes the SIFT descriptors for all the three color channels of the $O_1O_2O_3$ opponent color space, which is defined as:

$$\begin{bmatrix} O_1 \\ O_2 \\ O_3 \end{bmatrix} = \begin{bmatrix} \frac{R-G}{\sqrt{2}} \\ \frac{R+G-2B}{\sqrt{6}} \\ \frac{R+G+B}{\sqrt{3}} \end{bmatrix}. \quad (1)$$

The final descriptor length is thus $3 \times 128 = 384$.

The rg-SIFT descriptor computes the SIFT descriptor of the r and g chromaticity components of the normalized RGB color space, i.e.:

$$\begin{bmatrix} r \\ g \end{bmatrix} = \begin{bmatrix} \frac{R}{R+G+B} \\ \frac{G}{R+G+B} \end{bmatrix}, \quad (2)$$

and concatenates them with the original SIFT descriptor. The descriptor length is thus $3 \times 128 = 384$.

3.2 Hypothesis Validation

The hypothesis validation is divided into two different step: a geometric validation and appearance validation. In the geometric validation, the affine transformation A determined in step (7) is decomposed using QR-decomposition: $A = QR$, where Q is an orthogonal matrix ($Q^T Q = I$) and R is an upper triangular matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} s_x & s_h & t_x \\ 0 & s_y & t_y \\ 0 & 0 & 1 \end{bmatrix} = Q \cdot R. \quad (3)$$

The Q matrix gives information about the rotation angle θ . Since we do not pose any constraint about the admissible range of rotation, no validation is done on Q . On the other hand the matrix R gives information about scale, shear and shift. The geometric validation acts on R to exclude unfeasible transformations: first of all the shifts t_x and t_y must be in an admissible range (related to the image size) in order to be sure that the center of mass of the target is inside the image. The scales s_x and s_y cannot be too large, too small and too different from each other. The shear s_h cannot be too large. These conditions are given as thresholds: if an affine transformation passes all these conditions, then the appearance validation is performed.

In the appearance validation step the visual characteristics of the object we want to find are exploited. Since we know that the MCC is composed by 24 different colored patches and that the last row is composed of neutral patches, first of all the matrix A is inverted and applied to the image to obtain a normalized view of the object (i.e. the object aligned with the given template). The average

values of the R,G,B channels of the image at the expected patches position are extracted. The average values extracted from the last row are used to white balance the object. Then a 24-bins color histogram is extracted and compared with the one of the object template. The histogram is defined by a nearest neighbor quantization which uses as centroids the average RGB values of the 24 patches of the template.

The object pose with the most similar color histogram to the one of the template is then selected. The histograms are compared using the χ_2^2 similarity measure, which for two d -dimensional histograms H and H' is defined as

$$\chi_2^2(H, H') = \sum_{i=1}^d \frac{(H(i) - H'(i))^2}{H(i) + H'(i)}. \quad (4)$$

The whole hypothesis validation is synthesized in Figure 2.

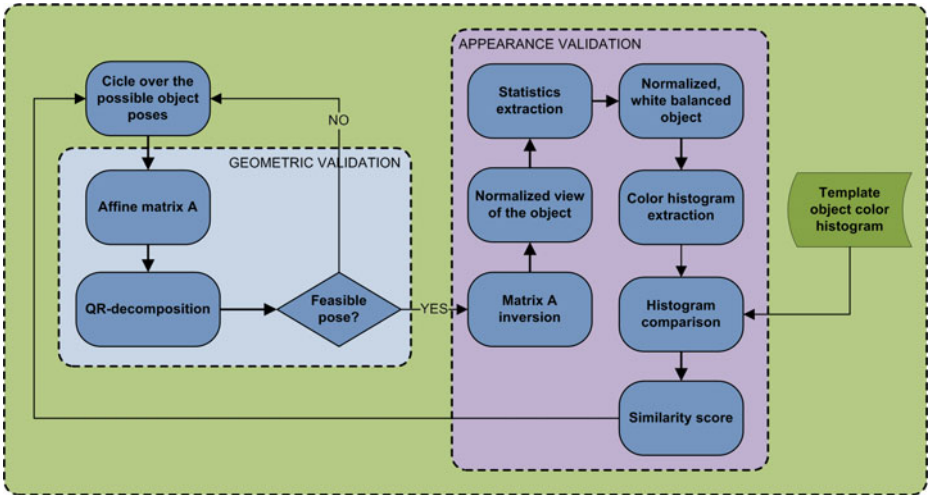


Fig. 2. Block diagram of the hypothesis validation. The two steps in which it is divided are clearly visible: the geometric validation and the appearance validation. In the geometric validation the localizations with a unfeasible target pose are discarded. In the appearance validation the object pose with the most visually similar object to the template is selected.

4 Experimental Results

To test the performance of the descriptor considered, a standard dataset of RAW camera images having a known color target has been used [5]. This dataset is captured using a high-quality digital SLR camera in RAW format, and is therefore free of any color correction. Using a freely available software (dcRAW, <http://www.cybercom.net/~dcoffin/dcraw/>) we have then demosaiced and

converted the images within the dataset into uncompressed linear 16-bit files, paying particular attention to convert the images always using the same multiplicative gains in order to bypass the camera Automatic White Balance (AWB) estimation. The dataset consists of a total of 568 images. The Macbeth ColorChecker (MCC) chart is included in every scene acquired and the task of the descriptors considered is to locate it. The performance measure used to identify the correctly located MCC color target is the same used for the detection task in the PASCAL Visual Object Class Challenge [13]. We consider a correct localization if the area of overlap a_o between the predicted bounding box B_p and the ground truth bounding box B_{gt} exceeds the threshold of 50%, where the area of overlap is computed as:

$$a_o = \frac{\text{area}(B_p \cap B_{gt})}{\text{area}(B_p \cup B_{gt})}. \quad (5)$$

The ratios of correct target localizations given by the descriptors considered are reported in Table 1. The most performing descriptor for the given task is the

Table 1. Ratios of correct localizations given by the descriptors considered

Descriptor name	Validation	Descriptor length	Correct localizations ($a_o > 50\%$)
SIFT	no	128	79.58%
Opponent-SIFT	no	$3 \times 128 = 384$	90.85%
rg-SIFT	no	$3 \times 128 = 384$	94.01%
SIFT	yes	128	89.79%
Opponent-SIFT	yes	$3 \times 128 = 384$	96.30%
rg-SIFT	yes	$3 \times 128 = 384$	96.30%

rg-SIFT with 94.01% of correct localizations. The hypothesis validation introduced was able to improve the ratio of correct localizations of all the descriptors considered: SIFT passed from 79.58% to 89.79%, opponent-SIFT from 90.85% to 96.30% and rg-SIFT from 94.01% to 96.30%.

In Figure 3 are reported the plots of the rates of correct target localizations as a function of the area of overlap a_o for the descriptors considered, both with and without the hypothesis validation.

In Figure 4 an example of the proposed approach is reported. In the first row the image in which we want to locate the color target and the multiple localizations extracted are reported. The second and third row contain two different localizations. From left to right, the first image is the normalized view of the localized area. The second one is the white balanced version of it. The third one is the 24-bin histogram extracted from the localized area and the last one is the 24-bin histogram of the reference target.

Examples of correct target localization using the rg-SIFT descriptor followed by the hypothesis validation are reported in Figure 5. Analyzing the localization

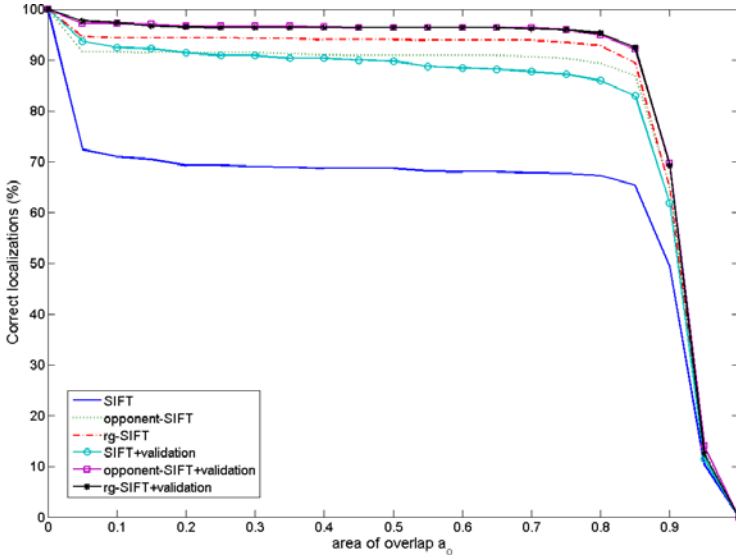


Fig. 3. Plot of the rates of correct target localization as a function of the area of overlap a_o for the descriptors considered, both with and without the hypothesis validation

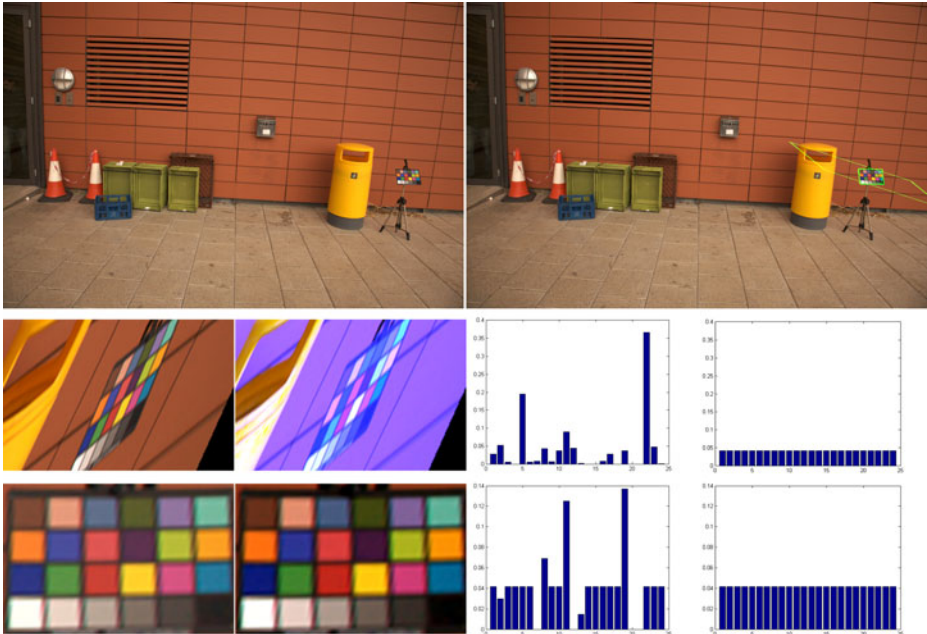


Fig. 4. Example of the proposed approach. First row: the image in which we want to locate the color target and the multiple localizations extracted. Second and third row: two different localizations. From left to right: normalized view of the localized area; white balanced version of it; 24-bin histogram extracted from the localized area; 24-bin histogram of the reference target.

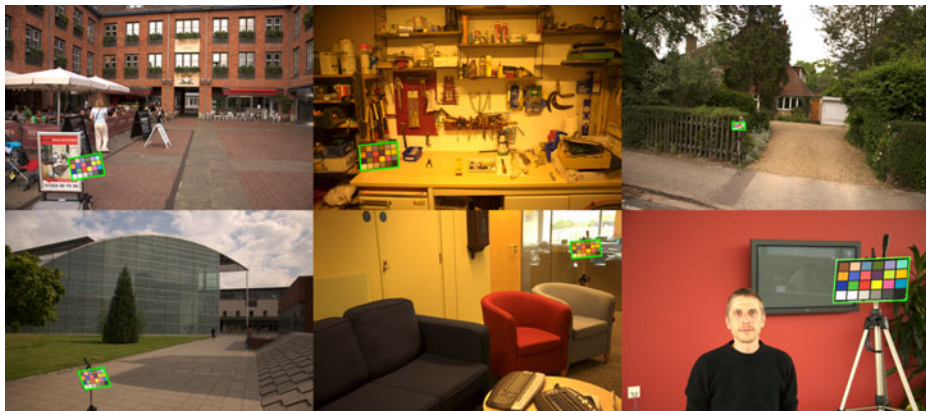


Fig. 5. Examples of correct target localization using the rg-SIFT descriptor followed by hypothesis validation

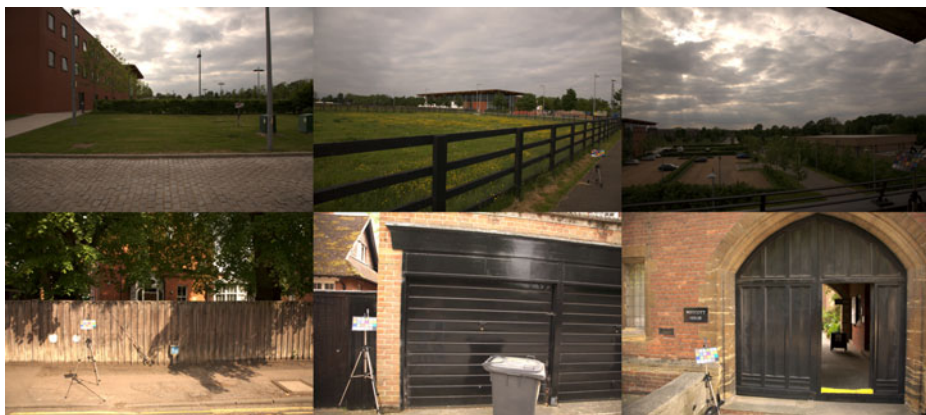


Fig. 6. Examples of images on which all the descriptors failed the target localization. The area of overlap a_o of the SIFT, opponent-SIFT and rg-SIFT, all followed by the hypothesis validation are (left to right, top to bottom): [0.07% 0% 0%], [9.84% 9.84% 0.16%], [0% 0% 0%], [0.24% 0.24% 0%], [3.15% 0% 0.58%], [8.93% 0.57% 0.57%]

results we noticed that among the images in the dataset there are 18 of them on which all the descriptors failed the target localization (considering an area of overlap $a_o > 50\%$). Examples of failed localizations are reported in Figure 6.

5 Conclusions

Given the importance of color targets in imaging applications and the need to use large datasets, in this work we have investigated the use of color descriptors

to automatically locate the color target in the scene. Three different gradient-based descriptors have been tested. These descriptors are then used to return multiple localization hypotheses and a geometrical and appearance validation are introduced to select the most feasible pose. The experimental results on a public dataset of RAW images containing the Macbeth ColorChecker CC target and acquired in uncontrolled environments, showed that the most performing descriptor for the given task is the rg-SIFT. The hypothesis validation introduced was able to improve the ratio of correct localizations of all the descriptors considered: a localization rate of more than 96% has been achieved with Opponent-SIFT and rg-SIFT descriptors.

Most of the missed targets were over or underexposed. As a future development, we plan to deal with these cases by performing a contrast enhancement before the appearance validation of candidate regions.

References

1. Martinez-Verdú, F., Pujol, J., Capilla, P.: Calculation of the Color Matching Functions of Digital Cameras from Their Complete Spectral Sensitivities. *Journal of Imaging Science and Technology* 46, 15–25 (2002)
2. Bianco, S., Gasparini, F., Russo, A., Schettini, R.: A new method for RGB to XYZ transformation based on pattern search optimization. *IEEE Transactions on Consumer Electronics* 53, 1020–1028 (2007)
3. Bianco, S., Gasparini, F., Schettini, R., Vanneschi, L.: Polynomial modeling and optimization for colorimetric characterization of scanners. *Journal of Electronic Imaging* 17(4), 043002-1–043002-13 (2008)
4. Hardeberg, J.Y., Schmitt, F., Brettel, H.: Multispectral color image capture using a liquid crystal tunable filter. *Optical Engineering* 41(10), 2532–2548 (2002)
5. Gehler, P.V., Rother, C., Blake, A., Minka, T., Sharp, T.: Bayesian Color Constancy Revisited. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008)*, pp. 1–8 (2008)
6. Smoyer, E.P.M., Taplin, L.A., Berns, R.S.: Experimental evaluation of museum case study digital camera systems. In: *Proceedings of the IS&T's 2005 Archiving Conference*, vol. 2, pp. 85–90 (2005)
7. Marguier, J., Bhatti, N., Baker, H., Harville, M., Süssstrunk, S.: Assessing human skin color from uncalibrated images. *International Journal of Imaging, Systems and Technology*, special issue on Applied Color Image Processing 17(3), 143–151 (2007)
8. Lowe, D.G.: Object recognition from local scale-invariant features. In: *Seventh International Conference on Computer Vision (ICCV 1999)*, vol. 2, pp. 1150–1157 (1999)
9. Harris, C., Stephens, M.: A combined corner and edge detector. In: *Proceedings of the 4th Alvey Vision Conference*, pp. 147–151 (1988)
10. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), 91–110 (2004)

11. van de Sande, K.E.A., Gevers, T., Snoek, C.G.M.: Evaluation of color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2009) (in press), doi:10.1109/TPAMI.2009.154
12. Bianco, S., Cusano, C.: Color target localizations using local descriptors. *Colore e Colorimetria: contributi multidisciplinari* (2010) (in press)
13. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL Visual Object Classes Challenge (VOC 2009) Results (2009), <http://www.pascal-network.org/challenges/VOC/voc2009/workshop/index.html>