Specularity, Shadow, and Occlusion Removal
for Planar Objects in Stereo Case

Irina Nurutdinova, Ronny Hänsch, Vincent Mühler, Stavroula Bourou, Alexandra I. Papadaki
and Olaf Hellwich

Computer Vision and Remote Sensing, Technische Universität Berlin, Marchstr. 23, MAR6-5, 10587 Berlin, Germany
{irina.nurutdinova, r.haensch, olaf.hellwich}@tu-berlin.de

Keywords: Specularity Removal, Shadow Removal, Occlusions, Planar Objects, Specular-free Image, Shadow-free Image, Occlusion-free Image.

Abstract: Specularities, shadows, and occlusions are phenomena that commonly occur in images and cause a loss of information. This paper addresses the task to detect and remove all these phenomena simultaneously in order to obtain a corrected image with all information visible and recognizable. The proposed (semi-)automatic algorithm utilizes two input images that depict a planar object. The images can be acquired without special equipment (such as flash systems) or restrictions on the spatial camera layout. Experiments were performed for various combinations of objects, phenomena occurring, and capturing conditions. The algorithm perfectly detects and removes specularities in all examined cases. Shadows and occlusions are satisfactorily detected and removed with minimal user intervention in the majority of the performed experiments.

1 INTRODUCTION

When photographing certain objects one is often faced with problems such as specularities, shadows, or occlusions. The properties of planar objects such as posters or books (i.e. being flat and reflective), the situation during image acquisition (e.g. crowded scenes during a poster presentation), as well as the fact that this type of images is usually taken to store information, make these problems especially severe in this use case. Figure 1(a) shows an example of all three phenomena and their effects on the image quality.

These effects commonly cause a loss of important information such as missing details in text and figures, are often beyond the control of the photographer, and can only marginally be resolved by postprocessing. However, they vary often with the viewpoint of the camera. While it might not be possible to find or use a viewpoint which allows to acquire images without these effects, different object parts will be corrupted if the images are taken from different positions.

This paper presents a practical solution to these problems by detecting and removing specularities, shadows, and occlusions for planar objects. The developed algorithm processes two input images taken from different views, captured by a consumer camera or a mobile phone. The proposed method does not require any professional equipment, such as high reso-

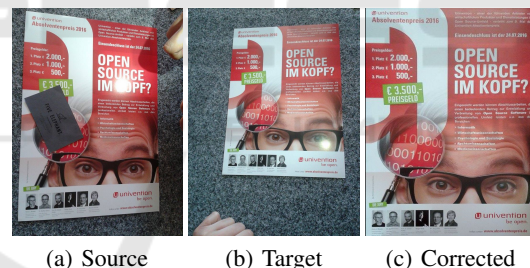

(a) Source   (b) Target   (c) Corrected

Figure 1: Specularities, shadows, and occlusions are automatically detected and removed leading to a visually pleasing result.

lution cameras, polarization filters, or flash structures. It does not rely on a specific spatial relation between camera and object, e.g. the distance to the object and the viewpoint of the camera, as long as the images of the object have a sufficient quality. The final result is a *corrected image* (see Figure 1(c) for the input images in Figures 1(a)-1(b)), that is free of specular areas, shadows, and occlusions, where the text is readable, the figures are complete, and the overall quality is close to that of a frontal, unoccluded picture.

The proposed method is implemented as a C++ desktop application and Android mobile app. It can be run automatically, but also allows interactive user input to improve results. The implementation is open source and freely available (Nurutdinova et al., 2016).

2 RELATED WORK

Although there are many methods aiming at the detection of specularities, shadows, or occlusions, the majority of those approaches deals with only one of these phenomena.

An overview of specularity removal techniques can be found in (Artusi et al., 2011). Methods for specularity detection can be divided into two groups based on whether they use a single or multiple images as input. The technique in (Klinker et al., 1988) analyzes the color space of a single image in order to estimate the distribution of specular pixels. The work in (Mallick et al., 2006) is based on a single image as well, analyzes the spatial neighborhood of a pixel, and uses partial differential equations to iteratively erode the specular components. The method of (Yang et al., 2013) separates specular and diffuse components in the HSI color space, clusters pixels with similar characteristics, and subsequently finds the optimal saturation.

Multi-image methods use multiple input images of the same scene but obtained from different points of view. The work in (Lin et al., 2002) uses color histogram differencing to retrieve specular pixels. It is based on color changes of specular regions among the different views which are estimated by computing the distance of the corresponding color histograms. The method for specularity removal of flat objects proposed in (Biasotti et al., 2015) employs a pixel value minimization across multi-view images which are captured by a mobile phone. It requires known camera orientation for each image, which is obtained by the built-in inertial measurement unit. The reference image is selected to be approximately parallel to the scanned surface.

A second category of approaches includes methods that require special equipment. The multi-flash method in (Feris et al., 2004) demands a flash system and uses an image sequence captured from fixed viewpoints with different positions of flash-light sources. The work in (Ma et al., 2007) is based on the fact that specular components of the reflected light are polarized. To separate these components, this method needs suitable polarization filters to measure the reflected light.

The detection of shadows is an important pre-processing step in many computer vision applications. The corresponding methods can be coarsely divided into pixel- and region-based approaches. An example of the former is the method in (Murali and Govindan, 2013) which detects shadows of a single image in the CIELab color space. The work in (Guo et al., 2011) is a region-based approach that classifies regions of the segmented image as shadow or non-shadow, according to their relative illumination.

Specularities and shadows can also be handled by deriving intrinsic images that decompose an image into its reflectance and illumination component. For example, the method in (Weiss, 2001) applies two derivative filters to a sequence of images to recover individual illumination images and a single reflectance map which is almost free of specularities and shadows. Several multi-view approaches for outdoor scenes (Laffont et al., 2013; Duchêne et al., 2015) compute a proxy geometry of the scene and utilize it for illumination estimation.

Occlusions can be considered as one of the biggest challenges in stereo vision. In (Zitnick and Kanade, 2000) the authors introduce a cooperative stereo algorithm. It is based on an uniqueness assumption and is jointly creating disparity maps as well as detecting occlusions. In order to detect occlusions in two-frame stereo, the iterative optimization algorithm in (Sun et al., 2005) uses a visibility constraint which is more general than the uniqueness constraint of (Zitnick and Kanade, 2000).

The technique proposed in this paper does not require any special equipment. Furthermore, two input images from two different viewpoints are sufficient while there are no specific constraints on the spatial relation between the two cameras and the object. In contrast to works such as (Biasotti et al., 2015), it is not necessary to determine the orientation of the cameras. Consequently, there is no predefined criterion for the selection of an image as the reference frame. The main contribution of the proposed approach is the creation of a *corrected image*, free of specularities, shadows, as well as occlusions for a planar object, even if the images contain all these undesired phenomena. The simplicity of the developed algorithm makes it robust and applicable in many use cases, i.e. it does not depend on specific image acquisition circumstances, high-end cameras, or other special equipment. Instead, images can be captured by hand-held consumer or mobile phone cameras from arbitrary viewpoints.

3 PROPOSED ALGORITHM

Specularities, shadows, and occlusions are effects that most commonly and most strongly affect the visibility of objects in images. They become especially severe in the case of images of planar and reflective surfaces which were acquired in crowded scenarios, e.g. posters in the interactive sessions of a conference. That is why the proposed technique focuses on the

joint detection and removal of these three phenomena for planar objects based on at least two images.

The goal of the algorithm is to create a *corrected image*, which is rectified and free of these phenomena by replacing corrupted regions with information from the input images. For this purpose, one image is selected as *target image T* that will provide the basis for the alignment, while affected regions are replaced by non-corrupted information from the other image, i.e. the *source image S*. An obvious requirement of this approach is that at least one of the input images needs to contain non-corrupted information for all object parts that are corrupted in the *target image*. In the case when a certain object area is corrupted in all input images, no correction can be performed unless more images of the object are taken.

This section presents the individual steps of the developed algorithm as shown in Figure 2, starting from a pair of input images until the creation of the *corrected image*.

3.1 Object Rectification

For the subsequent detection and correction steps (as described in Sections 3.3-3.5) it is sufficient to only align the given input images. However, potentially, none of the input images was captured with an image plane parallel to the object plane, leading to skewed projections of the object in all images. Furthermore, a significant part of the image might show the background, which might have a different depth than the object itself, leading to ambiguities during the automatic alignment (see Section 3.2).

That is why the proposed approach starts with an optional but recommended rectification step. The object is assumed to be rectangular (e.g. a poster or book) and its four corner points (see first row of Figure 2) are mapped to a predefined rectangular region by a first initial homography for source and target image, respectively. This step allows the user to specify the region of interest and increases the robustness, accuracy, and speed of the subsequent feature-based image alignment.

3.2 Image Alignment

The coarse initial alignment (Section 3.1) is not sufficiently accurate, if the corner points within the input images do not perfectly match each other. That is why this step computes a finer alignment based on keypoints detected by SIFT (Lowe, 2004) and matched among the images. If the optional object rectification step is left out (e.g. if one of the input images shows a sufficiently rectangular projection of the object with-
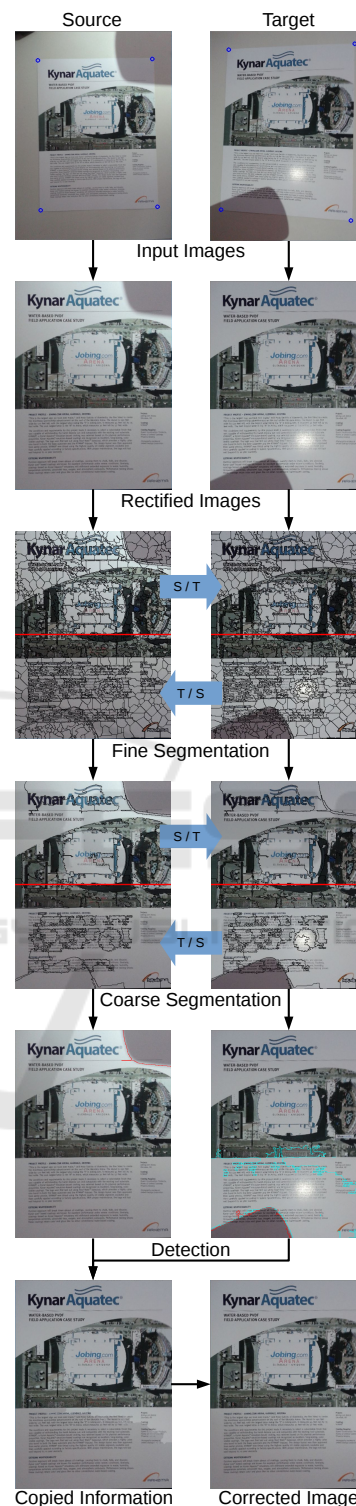

Figure 2: The pipeline of the developed technique, starting from two input images until the creation of the corrected image. The blue circles in the images of the first row denote the user-selected corners of the region of interest.

out too much background), this fine alignment is the first step of the whole processing chain.

RANSAC (Fischler and Bolles, 1981) is utilized to find the inliers of the obtained point correspondences, which are used for the calculation of the final homography which maps the (rectified) source image into the coordinate system of the (rectified) target image. Using this homography, the *source image S* is warped to the coordinate system of the *target image T*. Regions in the *target image*, that are not visible in the warped *source image*, are masked out in all subsequent steps. The second row of Figure 2 shows the result of the object rectification and alignment of the two input images shown in the first row.

3.3 Image Segmentation

Image segmentation is the next step of the proposed processing chain for several reasons. First of all, specularities as well as shadows and occlusion are regional phenomena that never affect single pixels but always whole parts of an image. Furthermore, methods based on isolated pixels are prone to incorrect classification due to image noise and imprecise image alignment. For this reason, the proposed method performs a classification that is based on image segments (e.g. superpixels) instead of single pixels by detecting color differences between segments of the *target* and *source image*. This does not only lead to an increased robustness and accuracy, but also to a significantly decreased computational load.

The two aligned input images are segmented into superpixels by SEEDS (van den Bergh et al., 2015), which is sufficiently fast, prevents over-segmentation of the image, and automatically determines the number of segments. In a second step the segment boundaries of the target image are projected into the source image leading to a target-source (T/S) segmentation and the segment boundaries of the source image are projected into the target image leading to a source-target (S/T) segmentation (an example is shown in the third row of Figure 2. This allows a valid comparison between image regions of the two images by avoiding the risk that two corresponding regions only have a mutual overlap but also contain disjoint parts of the image.

While the T/S segmentation is used to detect specularities, shadows, and occlusions in the *target image*, the S/T segmentation is used to detect shadow and occluded regions within the *source image*. It should be noted, that images without any deterioration would result in an equivalent S/T- and T/S segmentation (up to small variations due to noise). Significant differences in both segmentations are due to additional "objects" or features in the images such as specular regions, shadows, and occlusions.

This segmentation procedure leads to four different sets of superpixel $\left\{ s_i^{\alpha,\beta} \right\}_{i=1,...,N_\alpha}$, namely the S/T-segmentation derived from the source image ($\alpha = S/T, \beta = S$) and projected to the target image ($\alpha = S/T, \beta = T$), as well as the T/S-segmentation derived from the target image ($\alpha = T/S, \beta = T$) and projected to the source image ($\alpha = T/S, \beta = S$). For the sake of brevity the indices α, β are skipped, whenever they are not explicitly needed.

Each superpixel s_i is described by the mean color value of its pixels in CIELab space, denoted as $\mathbf{m}(s_i) = (m_L(s_i), m_a(s_i), m_b(s_i))$, where $m_c(s_i)$ with $c \in \{L, a, b\}$ is the mean value of the lightness and color-opponent channels, respectively.

While specularities often only affect rather small image regions that are sufficiently represented by superpixels, shadows and occlusions commonly cover large parts of the images. In order to obtain a larger spatial support and a more stable detection of these effects, superpixels are merged into larger image *regions*. The merging process is applied to superpixels $\left\{ s_i^{S/T,S} \right\}$ and $\left\{ s_i^{T/S,T} \right\}$ only, whereas the resulting regions $\left\{ r_u^{S/T,S} \right\}$ and $\left\{ r_v^{T/S,T} \right\}$ are projected into the other image to obtain regions $\left\{ r_u^{S/T,T} \right\}$ and $\left\{ r_v^{T/S,S} \right\}$.

The *regions* are formed by merging neighboring superpixels with similar colors. The region growing is initialized by defining each superpixel s_i as a region r_u of this image (where $\mathbf{m}(r_u)$ is the mean color vector of region r_u). An iterative process merges two adjacent regions r_u, r_v into one region if the condition in Equation (1) is fulfilled.

$$\|\mathbf{m}(r_u) - \mathbf{m}(r_v)\|_2 < m \tag{1}$$

The default values of the thresholds $m = 4$ has been determined empirically and is strict enough to preserve clear borders of most objects and shadows. Small *regions* with area below 5% of the image size are merged with the most similar adjacent *regions*, i.e. the ones with lowest difference of the mean color.

Figure 3(b) depicts the *regions* created by merging the superpixels of Figure 3(a).

3.4 Detection

In the detection step the previously obtained segmentation is used to classify image parts into specular regions, shadows, occluded, or not problematic regions. The classification is based on color differences between the segments within the aligned images.
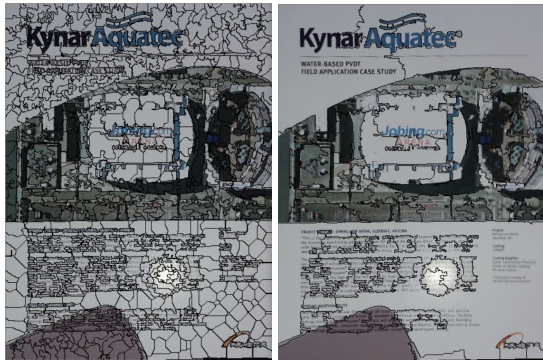
(a) Superpixels obtained by SEEDS

(b) Regions obtained by merging procedure

Figure 3: Result of image segmentation and merging.

While the shadow/occlusion detection (Section 3.4.1) operates on *regions*, the specularity detection (Section 3.4.2) uses the original *superpixels* (see Section 3.3).

3.4.1 Shadow and Occlusion Detection

Shadows and occluding objects usually appear as distinct, rather large areas with clear borders. Exploiting this characteristic, shadows and occlusions are detected as connected regions with sharp edges that are visible in one but not in the other image. For this reason the shadow and occlusion detection is based on the *regions* formed during the segmentation process as described in Section 3.3.

The neighborhood $N(r_u)$ is the set of all regions within the same image, that are adjacent to region r_u, i.e. share a common border. Let ∂r_u be the set of superpixels at the border of *region* r_u, i.e. each superpixel $s_i \in \partial r_u$ has at least one adjacent superpixel s_j which belongs to a neighboring region r_v, i.e. $s_j \in r_v$ with $r_v \in N(r_u)$. The exterior neighborhood $N_e(s_i)$ of a superpixel $s_i \in \partial r_u$ is the set of all adjacent superpixels which do not belong to the same region:

$$N_e(s_i) = \{s|s \in \partial r_v \text{ with } v \neq u \text{ and } r_v \in N(r_u)\} \quad (2)$$

For each superpixel $s_i \in \partial r_u$, the intensity difference to the superpixels in the exterior neighborhood $N_e(s_i)$ is computed and used to assign the corresponding label y to this superpixel by Equation (3).

$$y(s_i) = \begin{cases} 1, & \text{if } \exists s \in N_e(s_i) : m_L(s) - m_L(s_i) > t_{grad} \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

It should be noted that only object borders are marked where the interior superpixels are darker than the exterior.

After the object borders are defined according to Equation (3), they are compared among the different images. Each superpixel s classified as an object border in only one of the segmented images casts a vote to its underlying *region* r_u with $s_i \in r_u$ to be an additional object. The normalized sum of all votes in Equation (4) is considered as a cue for a shadow or occlusion.

$$Y^\alpha(r_u) = \frac{1}{|\partial r_u|} \sum_{s \in \partial r_u} (1 - \delta(y(s^{\alpha,S}), y(s^{\alpha,T}))) \quad (4)$$

where $\alpha \in \{S/T, T/S\}$ and $\delta(\cdot, \cdot)$ is the Kronecker delta function.

If a *region* r has sufficient number of votes, it is classified as *shadow* or *occluded region*. Specifically, if $Y(r_u) > a$ where $0 \leq a \leq 1$ is a constant.

A shadow/occluded region which is surrounded by other shadow/occluded regions might not obtain a sufficient number of votes, since the color difference between the corresponding boundary superpixels is too small. Thus, regions that share at least 80% of their border with other shadow/occluded regions are labeled as shadow/occluded as well.

3.4.2 Detection of Specularities

Specularities appear as regional phenomena usually smaller in size as occlusion and shadows. With the exception of strong and distinct highlights, the borders are often smooth and gradually change to the true intensity of the object. Typical examples are specularities such as highlights, reflected light flashes, mirror-like reflections, and overexposed areas.

The detection is based on the assumption, that a specular segment has a higher intensity value, since the color of the specular component adds to the underlying diffuse color. Therefore if Equation (5) holds, superpixel $s_i^{\alpha,T}$ is categorized as a specular candidate and is added to the *specularity mask* of the *source image*.

$$m_L\left(s_j^{T/S,T}\right) - m_L\left(s_j^{T/S,S}\right) > t_{spec} \quad (5)$$

The threshold t_{spec} is automatically determined based on the average intensity difference of all corresponding superpixels in *target* and *source* images.

Since shadows and dark occlusions in the *source image* appear brighter in the *target image* and thus fulfill Equation 5 as well, only superpixels are considered that have not been classified as shadow/occluded regions before. This effect is shown in Figure 4, where Figure 4(c) shows the result if this condition is not considered. Figure 4(d) shows the result if the previously performed shadow detection is exploited.

(a) Target image   (b) Source image
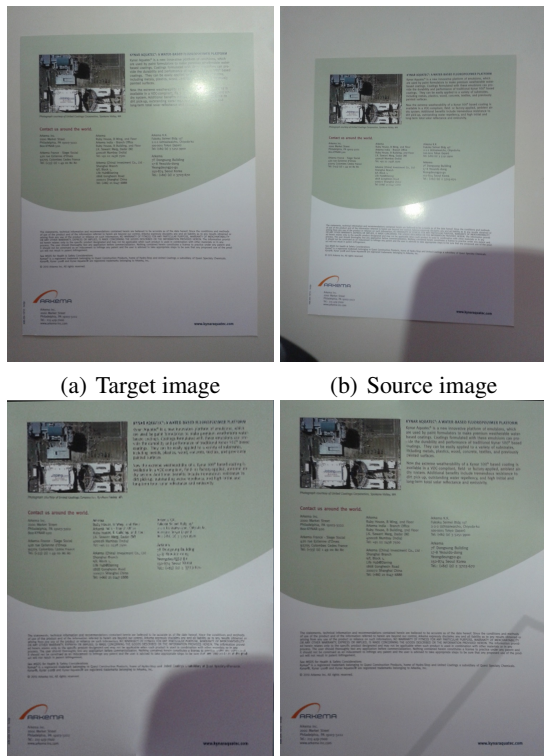
(c) Result using initial *specularity mask.*

(d) Result using final *specularity mask.*

Figure 4: Specularity detection.

(a) Target image   (b) Source image

(c) Result without Poisson Blending.
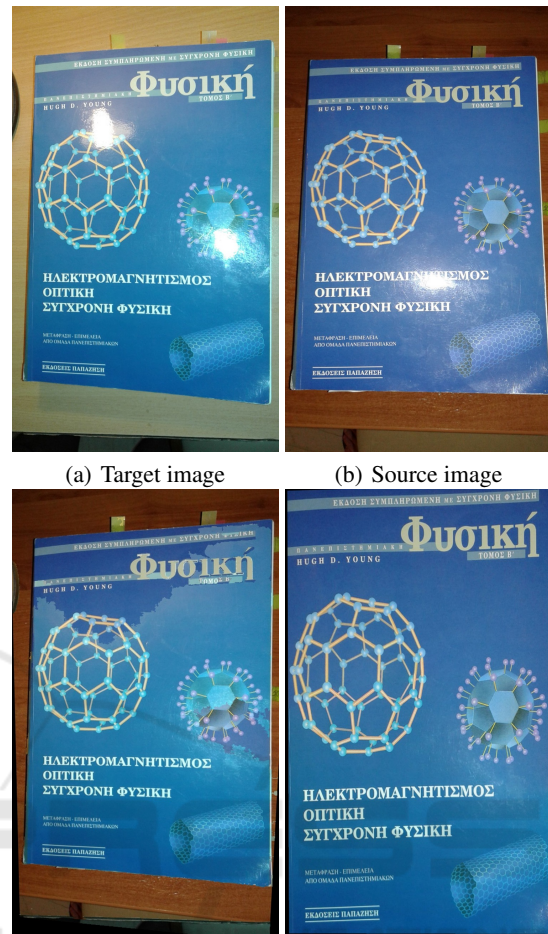
(d) Result with Poisson Blending.

Figure 5: Replacement of corrupted regions.

3.5 Image Correction

After all problematic regions in the source and target image have been defined and classified, they are used to obtain the corrected image which is free of those effects. In a first step, all detected problematic regions in the target image are combined to a replacement mask. There might occur small holes or small regions in this mask which are caused by fusing the different detections. They are filled or deleted, respectively, as a tradeoff between completeness of the replacement and visual consistent results: A very small shadow, occlusion, or specularity is unlikely to decrease the overall information content much, but small errors during the image registration might cause small misalignments which become most apparent for diagrams or text regions.

The entire area of the final mask is replaced in the *target image* by the corresponding area of the *source image*. The result is shown in Figure 5(c).

Due to different exposure and color balance of the two images, it is very likely that the replaced regions are clearly recognizable by color differences creating edges in the image which are not part of the object itself. Poisson Blending (Peréz et al., 2003) is used to seamlessly blend them with the regions of the target image. The main goal of this step is to obtain a visual consistent and pleasing corrected image, while fine details such as text in the blended areas are preserved. The final corrected image after Poisson Blending is shown in Figure 5(d).

4 USER INTERACTION

The proposed approach is often able to automatically detect and remove specularities, shadows, and occluded areas from the target image, provided that the corresponding regions are not degraded by these effects in the source image. However, the results of the method can be improved by minimal user interaction at three major points:

1. Object Rectification: As mentioned in Section 3.1, an object rectification can be performed which leads to an corrected image that shows only

the object of interest in a rectangular shape without any background. In this case, the user selects the four corner points of the object in the images which are then mapped to a predefined rectangle.

2. Shadow and Occlusion Detection: The distinct properties of specularities (i.e. small, very bright areas) make them easier to detect than shadows and occlusions, which can be very inhomogeneous and of (nearly) arbitrary shape.

   While the proposed method provides an automatic procedure, the user can optimize intermediate steps to achieve the better results. In a first step, it is possible to skip shadow/occlusion detection for individual images if the user determines them as not necessary (e.g. in the case that there is no shadow or occluded region). In this case, superpixels are not merged into *regions* for the *target-source segmentation* and the *source-target segmentation* is completely skipped if shadow and occlusion detection in the *source image* is not required. While this step has only a minor influence on the robustness of the result, it significantly decreases the computational complexity (and thus the run time) of the proposed method.

   Furthermore, the user can manually adjust the two parameters of the shadow/occlusion detection via track bars, such as the threshold t_{grad}, which describes the intensity of the shadow/object border and the number of votes needed for a *region* to be classified as a candidate *shadow region*.

   The voting segments as well as the detection result for the chosen parameter values is immediately displayed to the user to simplify the fine tuning.

3. Specularity Detection: Similar to the shadow/occlusion detection above, it is possible to adjust the threshold th_{spec}, which allows to increase or decrease the sensitivity to darker and brighter areas in the image.

5 RESULTS

Multiple experiments are performed using several images of planar objects in order to evaluate the algorithm. The tests are mainly focused on images of books and posters, which differ from each other in color, structure, content, as well as perspective distortions, and include image acquisition scenarios with different illumination conditions and camera baselines.

In total, 31 different image pairs have been used and are summarized in Table 1. We calculated the number of problematic regions such as specularities,
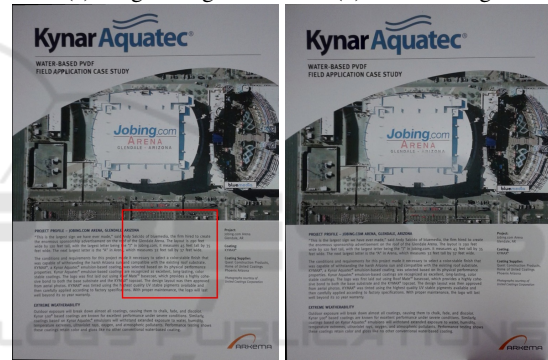
Table 1: Quantitative results of performed experiments.

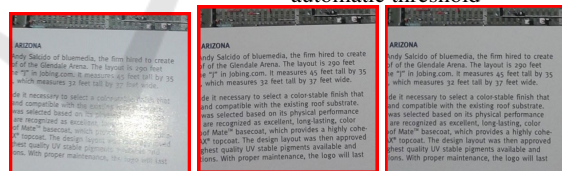|	T	S	C(auto)	C(manual)
Spec.	31	31	7	0
Shadows	13	6	3	1
Occlusions	14	10	6	1
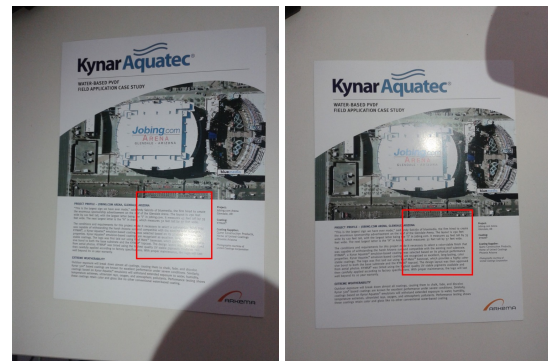
(a) Target image   (b) Source image

(c) Corrected image   (d) Corrected image with automatic threshold

(e) Detail in target image containing a specular region   (f) Detail in source image   (g) Detail in corrected image

Figure 6: Automatic and manual thresholds for shadow detection.

shadows and occlusions in *target*, *source* and *corrected images* (denoted as T, S and C respectively). All *target images* contained a specularity, while 13 of them additionally contained shadows. Occlusions occurred in 14 of the *target images*. None of the examined image pairs contained phenomena occurring in corresponding areas in both images.

The results of the experiments (summarized in Table 1) show that the algorithm successfully detects and removes the undesired phenomena for the ma-

jority of the examined cases. With fully automatic thresholds, we were able to remove 72% of all problematic regions. With user interaction, the problematic regions were detected in all but one of the examined cases.

The method is especially successful in the removal of specularities. Even in manual mode, the adjustment of the automatically calculated threshold for specularity detection was not usually required.

The proposed method is also successful in the detection of shadows. The most challenging phenomena are occlusions, because they may have very different properties, and it is hard to reason about the source of occlusion based on two views only. Figure 1 shows an extreme case where the target image contains all three kinds of degradations simultaneously. Nevertheless, all three problematic regions are successfully repaired, the shadow is removed, the title which was partially unreadable due to a strong highlight is restored, and the occluded region is filled by the correct information.

Another example is presented in Figure 6. Figure 6(c) shows the *corrected image* resulting from a manual adjustment of the thresholds by the user. On the other hand, Figure 6(d) depicts the result of the algorithm if the thresholds are automatically determined. The obtained corrected image is free of the specularity, while the shadow is not correctly detected and consequently not removed perfectly. Nevertheless, the major part of the obtained result is visually consistent. Figures 6(e)-6(g) show details of the target, source, and corrected image, respectively, denoted as red boxes in Figures 6(a)-6(c) and prove, that the readability of the text is preserved.

The most severe problems of the proposed method occur with shadows and occluding objects, when the occurring phenomenon has similar intensity values to the background. Such an example is shown in Figure 7, where the occluding object is detected and removed only in the area with different color in the background. Moreover, the shadow in Figure 7(a) is not detected, because it lacks a distinct border and rather presents a smooth intensity gradient in the image. As a consequence, both effects partially remain in the automatically computed corrected image (Figure 7(c)). If the thresholds are manually adjusted, the effects can be minimized but not fully corrected (Figure 7(d)).

6 CONCLUSIONS

This paper proposes an algorithm that successfully detects and removes specularities, shadows, and oc-

(a) Target image (b) Source image

(c) Corrected image with (d) Corrected image with
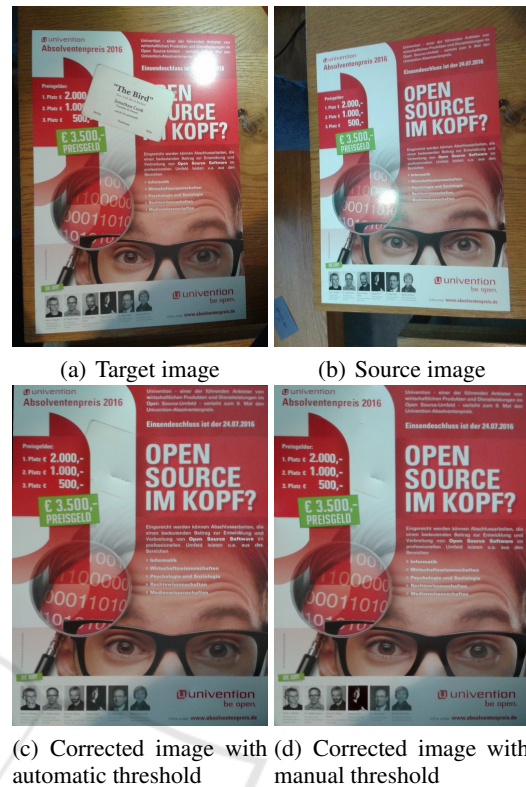automatic threshold   manual threshold

Figure 7: Case of undetected shadow and occlusion.

clusions using a two-frame technique for planar objects. Especially in the case of specular areas, the algorithm performs very well. However, the algorithm cannot detect shadows with very smooth borders or occlusions with a color similar to the object. Additionally, the algorithm can not remove phenomena that occur in the same object region in both images. Future work will fuse information from more images which increases the likelihood to find an image region non-corrupted information. The automation of the method should be increased by an robust procedure to select corresponding thresholds.

REFERENCES

Artusi, A., Banterle, F., and Chetverikov, D. (2011). A survey of specularity removal methods. In *Computer Graphics Forum*, volume 30, pages 2208–2230. Wiley Online Library.

Biasotti, S., Tarini, M., and Giachetti, A. (2015). Mobile multiview diffuse texture extraction. In *Smart Tools and Applications in Computer Graphics*, pages 113–120. Eurographics Association.

Duchêne, S., Riant, C., Chaurasia, G., Lopez-Moreno, J., Laffont, P.-Y., Popov, S., Bousseau, A., and Drettakis, G. (2015). Multi-view intrinsic images of outdoors

scenes with an application to relighting. *ACM Transactions on Graphics*, page 16.

Feris, R., Raskar, R., Tan, K. H., and Turk, M. (2004). Specular reflection reduction with multi-flash imaging. In *Computer Graphics and Image Processing, Proceedings of the 17th Brazilian Symposium*, pages 316–321. IEEE.

Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In *Communications of the ACM*, volume 24(6), pages 381–395.

Guo, R., Dai, Q., and Hoiem, D. (2011). Single-image shadow detection and removal using paired regions. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference*, pages 2033–2040. IEEE.

Klinker, G. J., Shafer, S. A., and Kanade, T. (1988). The measurement of highlights in color images. *International Journal of Computer Vision*, 2(1):7–32.

Laffont, P.-Y., Bousseau, A., and Drettakis, G. (2013). Rich intrinsic image decomposition of outdoor scenes from multiple views. *IEEE transactions on visualization and computer graphics*, 19(2):210–224.

Lin, S., Li, Y., Kang, S. B., Tong, X., and Shum, H. Y. (2002). Diffuse-specular separation and depth recovery from image sequences. In *European conference on computer vision*, pages 210–224. Springer Berlin Heidelberg.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110.

Ma, W. C., Hawkins, T., Peers, P., Chabert, C. F., Weiss, M., and Debevec, P. (2007). Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In *In Proceedings of the 18th Eurographics conference on Rendering Techniques*, pages 183–194. Eurographics Association.

Mallick, S. P., Zickler, T., Belhumeur, P. N., and Kriegman, D. J. (2006). Specularity removal in images and videos: A pde approach. In *European Conference on Computer Vision*, pages 550–563. Springer Berlin Heidelberg.

Murali, S. and Govindan, V. K. (2013). Shadow detection and removal from a single image using lab color space. *Cybernetics and information technologies*, 13(1):95–103.

Nurutdinova, I., Hänsch, R., Mühler, V., Bourou, S., and Papadaki, A. I. (2016). Project Website. `http://www.rhaensch.de/specular.html`.

Peréz, P., Gangnet, M., and Blake, A. (2003). Poisson image editing. In *ACM Transactions on Graphics (TOG)*, volume 22(3), pages 313–318. ACM.

Sun, J., Li, Y., Kang, S. B., and Shum, H. Y. (2005). Symmetric stereo matching for occlusion handling. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 399–406. IEEE.

van den Bergh, M., Boix, X., Roig, G., and van Gool, L. (2015). Seeds: Superpixels extracted via energy-driven sampling. *International Journal of Computer Vision (IJCV)*, 111(3):298–314.

Weiss, Y. (2001). Deriving intrinsic images from image sequences. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 68–75. IEEE.

Yang, J., Liu, L., and Li, S. (2013). Separating specular and diffuse reflection components in the hsi color space. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 891–898.

Zitnick, C. L. and Kanade, T. (2000). A cooperative algorithm for stereo matching and occlusion detection. *IEEE Transactions on pattern analysis and machine intelligence*, 22(7):675–684.