Specular object detection using combination of image and motion cues

Anonymous ECCV submission

Paper ID 690

Abstract. Identification and localization of shiny objects is an important task for a vision system in order to extract the semantic content of the images, and effective interaction with the environment. Previously, an algorithm was proposed that exploits motion cues in videos for shiny object detection and localization. In this paper we are extending the algorithm by fusing motion and static image cues which dramatically improves detection performance. Motion cues are related with epipolar deviations and appearance distortions due to specularity, while image cues are extracted via sliding window based object classification. We test the effectiveness of the proposed algorithm on various object shapes, object sizes, surface properties and noise/blur conditions. The proposed algorithm provides a complete solution to specular object detection, successfully localizing shiny objects in a wide variety of conditions.

Keywords: Specularity detection, Image motion, Object classification, Surface reflectance estimation

1 Introduction

One of the fundamental problems in computer vision is surface reflectance estimation. Surface reflectance is a prerequisite for the successful recovery of 3D shape [1–3], and it also provides crucial information about the semantic identity of objects (Figure 1). For an artificial agent visual extraction of reflectance properties may be a crucial prerequisite for properly planning interactions with the environment, for example in the biomedical context [4], or in real-time , real world 3D reconstruction scenarios [5].

The object's reflectance properties, its 3D shape and the illumination have to be estimated simultaneously from the 2D patterns of light arriving at the sensor therefore visual estimation of surface reflectance properties is mathematically under-constrained. Specific assumptions about the spectral BRDF [6, 7] or knowledge of camera motion [8, 9] has been made under specific conditions in the majority of previous work on reflectance classification, such as specialized sensing or lighting [10].

Image motion has been beneficial in many computer vision problems: structure from motion, image stitching, 3D shape recovery, stereo correspondence, recognition, or pose estimation, and has recently received increasing attention

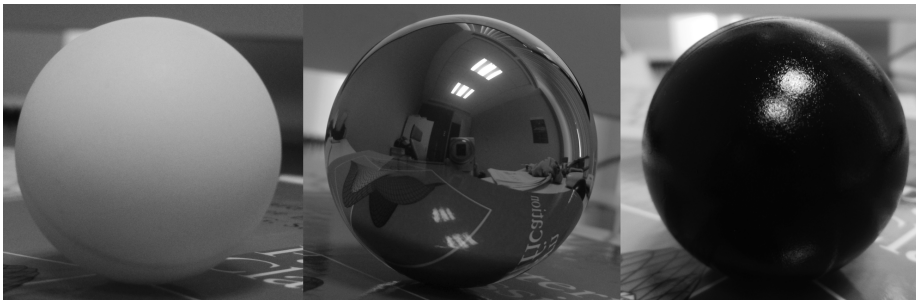

Fig. 1. Surface reflectance, appearance and identity. The shape in these three photographs is the same but the identity of the object changes as a function of its surface reflectance characteristics. Semantic labeling would be impossible on the basis of shape alone. From left to right: pingpong ball, chrome ball bearing, black plastic sphere. To optimally interact with these objects, e.g. to pick them up without breaking or dropping them, visual estimation of surface reflectance is crucial

in 3D specular shape reconstruction [11], specularity detection [12, 13], and reflectance classification [14].

The evidence in [15] suggests that one particular powerful image motion cue that the human visual system seems to be sensitive to when estimating surface reflectance is the distortion of appearance, that moving specular surfaces give rise to [14, 15]. Appearance distortion is an attractive motion feature to be extracted by a specularity detecting algorithm because it does not require any assumptions about the object, the illumination or the camera trajectory, it can be computed from just two images of a motion sequence, and it is a robust feature: it consistently occurs on specular surfaces, regardless of object shape, motion, or the reflected environment.

Another motion cue for specularity detection is epipolar deviation. For rigid, diffusely reflecting objects the optic flow due to camera motion obeys epipolar geometry. [12] showed that motion of specular object violates the epipolar constraint. While these epipolar deviations may signal the existence of specularities [12, 13], their usefulness in specular object detection is restricted to linear camera motion and convex shapes. Moreover, problems may arise - due to the global nature of this cue (at the fundamental matrix estimation stage, motion vectors from the entire image contribute to computation) - that diminish its specificity. For example, moving matte, textured objects will distort the optic flow field solely due to camera motion, and would thus cause epipolar deviations. Moreover, slowly moving, near planar, specular objects will have negligible epipolar deviations, and may thus not be identified as specular [14]).

An algorithm developed in [15] relies on scale and rotation invariant feature extraction techniques and uses the two motion cues (appearance distortion and epipolar deviation) to detect and localize specular surfaces in both computer-rendered and real image sequences. Change in feature vectors due to appearance change is used to quantify the appearance distortion on specular surfaces. The

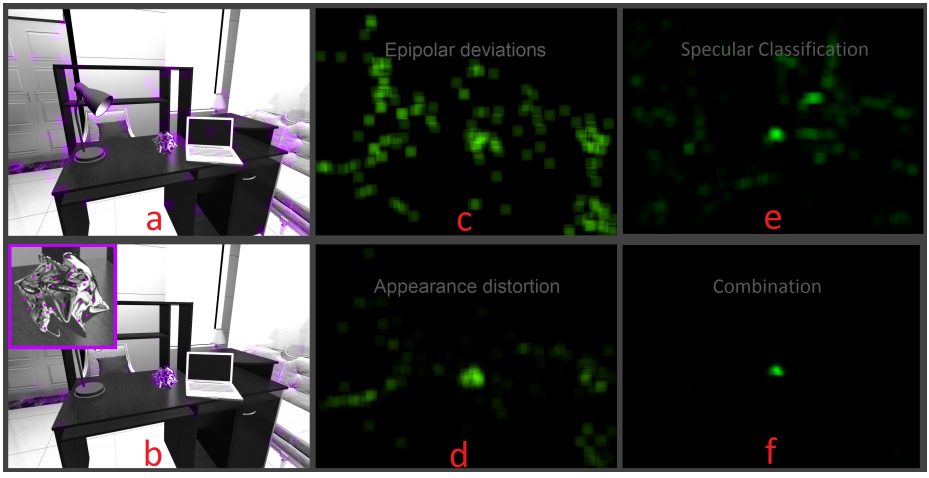

Fig. 2. Sample Scene. The camera motion is a rotation in azimuth, a single specular object is located on the desk. (Best viewed in color).
(a) A sample frame from the image sequences we used in our experiments. The purple color indicates epipolar deviation cue based detections
(b) The specular object is shown in close shot and the purple color indicates appearance distortion.
(c) Epipolar deviation (outlier) field shown separately.
(d) Appearance distortion field shown separately.
(e) Specular classification field obtained by sliding window based specular object search.
(f) The combined specularity field. Inset figure in b illustrates that appearance distortion, and therefore or algorithm also detects concave specular regions.

algorithm combined appearance distortion cue and epipolar deviations [12]. Even though combination of the two motion cues successfully detects shiny objects in videos, the algorithms have its shortcomings: the two motion cues are unable to detect low curvature shiny objects or objects with low reflectance. In this paper we are combining motion cues with image cues (Figure 2), and the proposed algorithm is performing dramatically better than motion-only algorithm in [15], not only in low curvature or reflectance conditions but in all scenarios. We are reporting five-fold improvement in detection precision, for failed scenarios in [14] and [15]. Image based specularity detection utilizes local binary pattern based feature extraction and linear SVM classifier for sliding windows based detection. Overall, we are providing a complete specularity detection system which works in a wide variety of object shapes and surface properties.

The paper is organized as follows: in section 2 we introduce the algorithm for detection and localization of specularities in image sequences. Section 3 describes the test sets and in section 4 we present the experimental results for various camera motions, camera speeds, object sizes, object shapes, surface reflectance properties in computer- generated scenes, as well as for videos. We end with a brief discussion in section 5.

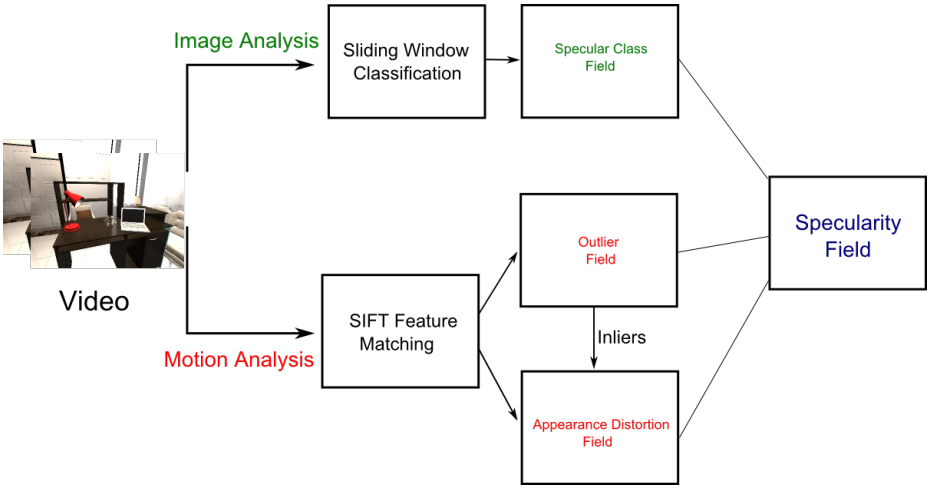

Fig. 3. The algorithmic flow of specular object detection. It has two main branches: Image and Motion analysis. Image analysis outputs specular class field. Motion analysis relies on two separate pathways: outlier and appearance distortion fields. The three fields are combined to obtain the specularity field.

2 Algorithm

The specular region detection algorithm that we propose takes a pair of images (current and previous) and outputs a specularity map in the current image (Figure 2 and Figure 3). The algorithm has two distinct stages: Image and Motion Analysis. In Motion Analysis we use SIFT features [16, 17]. They not only provide scale and rotation invariance in image analysis but also feature vectors that quantify local appearance. This makes them a perfect tool to directly capture the appearance distortions that occur on specular surfaces. In Image Analysis we train a classifier that separates specular objects from non-specular objects. At the end we combine three cues for obtaining specularity field of the image (see Figure 2). We give the details of the three cues in the following subsections (used parameters are given in the Appendix).

2.1 Epipolar Deviations

For computing epiploar deviations we utilize SIFT feature matching of consecutive frames in the video, then use fundamental matrix estimation to discriminate outlier matching. The flow is given as follows:

1. Extract SIFT features for each frame.
2. Eliminate features with low average feature vectors (Appendix)
3. SIFT nearest neighbour feature matching [18].
4. Apply 2000 RANSAC [19] iterations with 8 point DLT fundamental matrix estimation [20] to matching features.

5. Accept features with Sampson error [21] more than a selected threshold as outliers (Appendix).
6. Initializing a zero magnitude field (same size as the image).
7. Assign high intensity (value 1) to outlier pixels and convolve with Gaussian kernel (Appendix).

This field quantifies the density of epipolar deviations in the image. See Figure 2.

2.2 Appearance Distortion

For computing appearance distortion we also use SIFT features. The feature vector is a robust representation of the local appearance on objects. We measure the change in matching feature vectors to quantify appearance distortion. The flow is given as follows:

1. Compute L1 norm of the change in matching feature vectors for consecutive frames (inliers only). L1 norm is chosen over L2 norm, for its relative sensitivity to smaller distances.
2. Initializing a zero magnitude field (same size as the image).
3. Assign appearance distortion values (difference in matching vectors) to inlier pixel locations and convolve with the same Gaussian kernel as in section 2.1.

The resulting field quantifies the appearance distortion in the image (see Figure 2). Note that, appearance changes for outlier features are not used in order to avoid false high appearance distortion values. The latter arise from errors in feature matching.

2.3 Sliding Window Based Classification

In order to detect specular regions in images, we need to train a classifier and a sliding window based search procedure. We used Flicker Material Database (FMD) [22] as our source of specular/non-specular images. (Detailed information about FMD dataset is given in our supplementary document.) We reshaped the size of all images to 512 by 384 pixels, and then extracted local binary pattern (LBP) features [23, 24]; (58 quantized implementation of Vedaldi [17]). The cell size of the LBP was 128 pixels. The LBP features of the images were used for training multi class linear SVM classifier with L1 norm.

It should be noted that a relatively small (100 images per class, total of 10 classes) and a completely independent dataset was used for training the specular classifier. We selected glass subset of the FMD dataset as specular class and the rest of the images as non-specular class. During test time, images were searched for specular image blocks (specifically images that are classified as glass according to FMD) using sliding windows of size 64 by 48 pixels, and a skip size of 5 pixels. Centre pixel of the sliding window represents a detected specular object. We initialized a specular classification field with zero values, assigned a weight value to specular pixels and convolve with Gaussian kernel (Appendix).

2.4 Combination of Cues

The specularity field is obtained through pixel by pixel multiplication of the epipolar deviation, appearance distortion and specular class fields (see Figure 2 and 4). Multiplication of three fields acts as an AND operation, but it is more quantitative and flexible since a threshold can be defined on the combined field. This field has high intensity only if both epipolar deviations, appearance distortion exists and also the probability that class label of the pixel being specular is high. The combined field is thresholded to declare the final detected specular pixels.

2.5 Performance evaluation

We used precision and recall metrics for evaluating the detection performance of our system. We first determined the intersection of the detected specular pixels with the area delimited by the ground truth bounding box. Precision is the ratio of the number of intersection pixels to the number of detected specular pixels. Recall is the ratio of the number of intersection pixels to the number of ground truth specular pixels. Precision is equal to one if the detected specular pixels are completely inside the ground truth bounding box. It should be as large as possible in order not to declare false specular pixels. If recall is larger than a satisfactory threshold (eg.0.2), then enough number of specular pixels is retrieved to declare specularity detection. Additionally the system needs to avoid false detection, hence give a satisfactory precision. For this purpose, we will primarily use average precision to evaluate the performance of the algorithm. In the experiments we will report the average precision in the Precision-Recall curve, for the cases when Recall is larger than 0.2. Please note that, the experimental findings are not sensitive to the choice of recall threshold (See supplementary material for more results).

3 Test Set

We used the same dataset in [15] in order to demonstrate the effect of cue combination. For the details of the dataset generation, readers are invited to read section 4 of the paper. The dataset consists of carefully designed rendered video sequences that cover various object shapes, camera motion, object reflectance property, noise and blur levels; as well as real video sequences. The explanation of the dataset images are given in the experiments section. Matlab code of experiments are available at our website (`Anonymous_due_to_blind_review`).

4 Experimental Results

4.1 Effect of Frame Integration

Integration is one of the main methods in a wide spectrum of scientific fields, to improve signal to noise ratio. In the context of video based detections, collecting the result for multiple frames and combining them is a primitive and fast

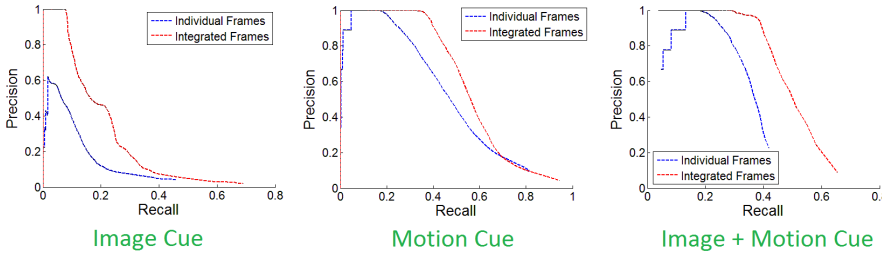

Fig. 4. The effect of frame integration is demonstrated for three types of cues (Image, Motion and Image+Motion). Precision recall curves are given for two conditions. Blue curve is individual frame based detection and red curve is frame integration based detection. Only image cue is used (left), only motion cue is used (middle) or motion and image cues are combined (right).

Table 1. The effect of object size and type of camera motion on average detection precision is given. (L-Large; R-Rotation; Z-Zoom; S-Small; T-Translation)

(Object Size, Camera Motion)	Image Cue	Motion Cue	Image + Motion Cue
(L, R)	0.20	0.76	0.96
(L, T)	0.13	0.77	0.87
(L, Z)	0.10	0.66	0.89
(S, R)	0.18	0.80	0.96
(S, T)	0.05	0.75	0.87
(S, Z)	0.07	0.67	0.87

kind of frame integration. We adopted this approach to improve the precision of specular object detection. For this purpose, we added the specularity fields of consecutive frames (10) and applied thresholding to the integrated frame instead of individual frames (integrated the last 9 frames with the current). In Figure 4, we show that integration significantly improves precision of detection, for both image (left) and motion (middle) cues as well as the combined cues (right). The rest of the experimental results reported in the paper uses frame integration.

4.2 Combination of Motion and Image Cues

As previously explained, motion cues have shortcomings in detecting spherical surfaces or surfaces with low reflectance [15]. In this paper, we introduce classification based image cue to support motion cues discovered in [15]. Figure 5 reports the performance of motion and motion+image cues for three different conditions. Although image cue is inferior in performance compared to motion cue, when the two cues are combined, the detector becomes dramatically better. This observation implies that motion and image cues collect complementary information.

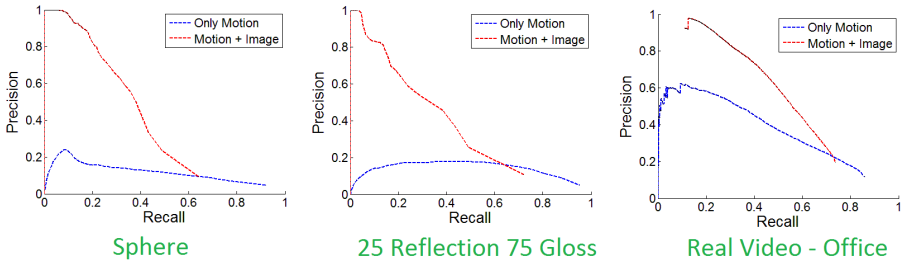

Fig. 5. The effect of image motion cue integration is demonstrated for various videos, previously reported to be problematic for motion cue based detection. Precision-Recall curves are given for Sphere shaped object (Left), Low reflectance high glossy surface (Middle) and real video sequence. In all three cases image and motion cue combination signficantly improves detection performance compared to motion cue alone.

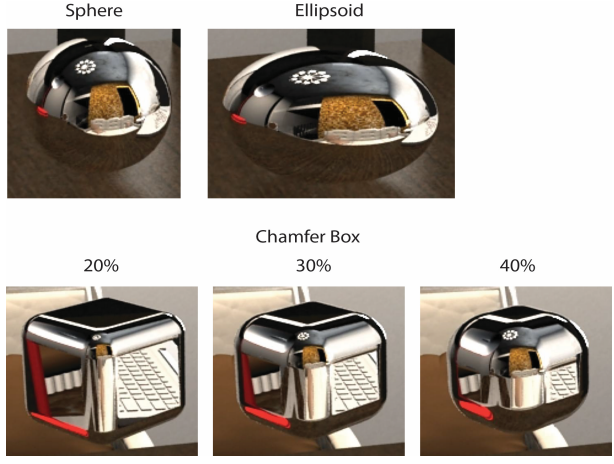

Fig. 6. Low curvature objects previously reported to be problematic for motion cue based detection are shown. Sphere and ellipsoid has very smooth surface. The smoothness of a cuboid edge is controlled by chamfer parameter (larger the smoother).

We also examined average detection precision for various object size and camera motion scenarios (Table 1). There is no significant effect of size and camera motion types for our experiments, however these set of experiments also verify the fact that image and motion cue combination is extremely beneficial for detection, for various size and motion scenarios.

4.3 Object Shape

Motion cue was shown to be unable to detect specularities on spherical objects, or objects with smoother undulations (see Figure 6) [15]. We tested detection

Table 2. The effect of object shape on average detection precision is given.

Shape	Image Cue	Motion Cue	Image + Motion Cue
Sphere	0.07	0.12	0.60
Ellipsoid	0.07	0.11	0.33
Chamfer 20%	0.05	0.28	0.60
Chamfer 30%	0.07	0.31	0.63
Chamfer 40%	0.09	0.30	0.87

performance for a set of simple objects that varied in their surface curvature: a sphere, ellipsoid and three different chamfer boxes with different curvatures (Figure 6).

We show in Table 2 that, combination of image and motion cues shows the most impressive improvements in detection, with 5 fold increase in precision for sphere object. Again, the image cue provides a very poor detector, yet it dramatically aids the motion cue to give a much superior performance.

4.4 Reflectance Properties

Similar to object shape, studies in [15] reported a poor performance for objects with low reflectance surface (Figure 7). The cue combination is a remedy to that problem as we report significant improvements in Table 3. In this table we report average precision for various reflectance and glossiness conditions represented by (Reflectance,Glossiness) duplet. It is observed that combination of cues enhances detection performance for every combination of surface properties. The highest improvement, 4-fold increase in average precision, is observed for the worst glossiness condition, (1, 0.5).

Table 3. The effect of reflectance and glossiness on average detection precision is given.

(Reflectance, Glossiness)	Image Cue	Motion Cue	Image + Motion Cue
(1, 1)	0.2	0.76	0.96
(1, 0.75)	0.03	0.77	0.81
(1, 0.5)	0.13	0.22	0.80
(0.75, 1)	0.07	0.77	0.93
(0.75, 0.75)	0.04	0.59	0.85
(0.5, 1)	0.18	0.64	0.90
(0.25, 1)	0.1	0.26	0.78
(0.25, 0.75)	0.05	0.15	0.43

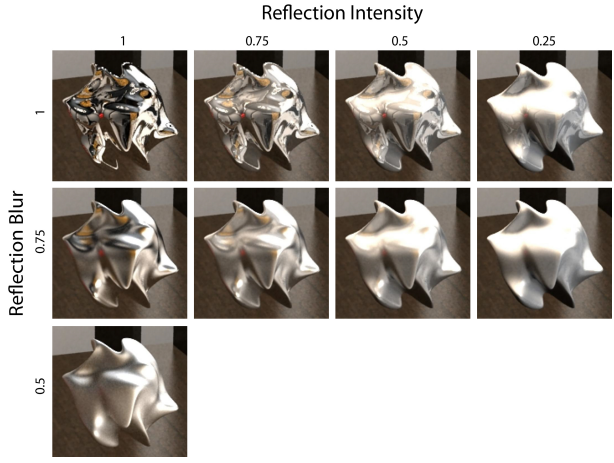

Fig. 7. The effect of varying reflection intensity and blur on specularity is illustrated. Previous motion based alglorithm was reported to give poor detection performance for low reflection intensity and high blur conditions. The algorithm is tested on all the objects to evaluate the sensitivity to surface properties.

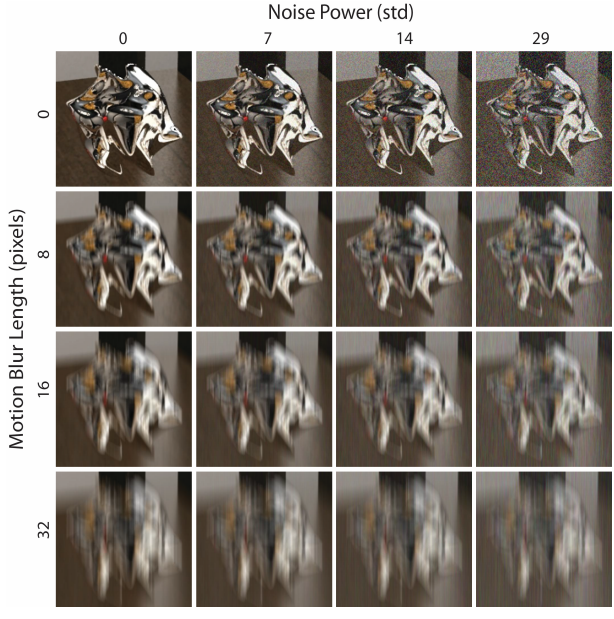

Fig. 8. The effect of noise and blur on specular object is demonstrated for various noise and blur levels is given.

Table 4. The effect of noise and blur on average detection precision is given.

(Noise, Blur)	Image Cue	Motion Cue	Image + Motion Cue
(0, 8)	0.04	0.91	0.84
(0, 16)	0.03	0.83	0.85
(0, 32)	0.02	0.86	0.74
(7, 0)	0.12	0.82	0.86
(7, 8)	0.04	0.90	0.88
(7, 16)	0.04	0.82	0.89
(7, 32)	0.03	0.89	0.81
(14, 0)	0.07	0.87	0.87
(14, 8)	0.03	0.92	0.78
(14, 16)	0.04	0.86	0.87
(14, 32)	0.03	0.85	0.83
(29, 0)	0.05	0.91	0.90
(29, 8)	0.02	0.93	0.89
(29, 16)	0.03	0.88	0.84
(29, 32)	0.03	0.81	0.71

Table 5. The performance of the detectors on real video sequence is given.

Image Cue	Motion Cue	Image + Motion Cue
0.13	0.39	0.75

4.5 Noise and Blur

We examined the effect of image noise and blur (Figure 8) on detection performance. We introduced additive Gaussian noise (σ , 0–29) and motion blur (length, 0–32 pixels) to our image sequences, varying noise power and motion blur length (Figure 8). Noise was added to the image before applying motion blur. This is a harder case than the opposite sequence since it creates structured noise in the image.

The results are reported in Table 4, in which we show that the detection system is capable of specularity detection for very wild scenarios of noise and blur. Yet, for some cases, cue combination slightly decreases the performance due to the fact that image cue is much more sensitive to noise and blur than epipolar deviation cue.

4.6 Real Video Sequence

For real-world experiments we used the sequences in [15]. The algorithm had to find this object in a real-world scene Office. Note, that this movie was taken under everyday conditions (handheld camera, accidental trajectories, etc.), thus should constitute a hardest-case test scenario (Figure 9). To make our tests

highly stringent we left the parameters of the algorithm **unchanged** [1], i.e., they were optimized for performance in computer-rendered scenes. We report in Table 5 that the combination of cues significantly improves detection performance for real video sequences. Thus the proposed algorithm is proven to be useful for real videos as well as rendered videos.

5 Discussion

Surface reflectance is a major factor contributing to an object's appearance, and estimation of surface reflectance is a fundamental problem in computer vision. Recently, image motion has been shown to provide useful information for reflectance classification [14] and specularity detection [12, 13]. [15] proposed a robust algorithm for specularity detection in videos, using a set of motion cues. Here, we developed a novel algorithm for the detection and localization of specular objects using image motion, that combines motion cues with classification based image cues.

Image based detection was achieved by training a specular object classifier on a publicly available material image dataset (FMD), and performing a sliding window based search over the test images for specularities. The training dataset is quite small (100 images per class and only glass class is used to detect specular objects), yet the trained classifier performs surprisingly well. More importantly, the image based detection aids the motion cue dramatically and together the two cues build a very robust and successful detector.

We tested the algorithm under a wide range of conditions. Cue combination is shown to be extremely beneficial for detection, and image cue is complementary to motion cue. Motion cue measure how the specular surface appearance abruptly with motion and how it does not obey 3D motion of the frame, whereas the image cue captures the general appearance of the specular surfaces in static images. Hence they are complementary in nature due to the added dimension of time in motion cue.

Performance was excellent for all types of camera trajectories (translation, zoom, rotation). The results on very different object variations, noise, blur etc. showed the generalization power of the algorithm.

The proposed algorithm provides a complete solution to specular object detection, successfully localizing shiny objects in a wide variety of conditions. Yet, there is still room for improvement, possibly using a larger dataset for classifier training, a better feature extraction routine, and using a more capable classification algorithm.

It is observed that the proposed algorithm performs poorly when the amount of noise and motion blur increases. This is due to the fact that image based specularity detection is impaired significantly with noise and blur.

[1] There were 3 real video sequences in [15] but 2 of them were not appropriate due to frame integration. In order to test on them, the algorithm should be modified, and we prefer not to do that due to consistency

Even though the main focus of our study is not developing an image based specularity detection algorithm, as a future work we would like to compare our image based detection algorithm with existing methods [25][2]. Specularity detection functionality can be imported to robotic vision applications for slippery surface detection, rendering of 3D reconstructed specular surfaces or to production line vision applications for detecting shiny metal objects.

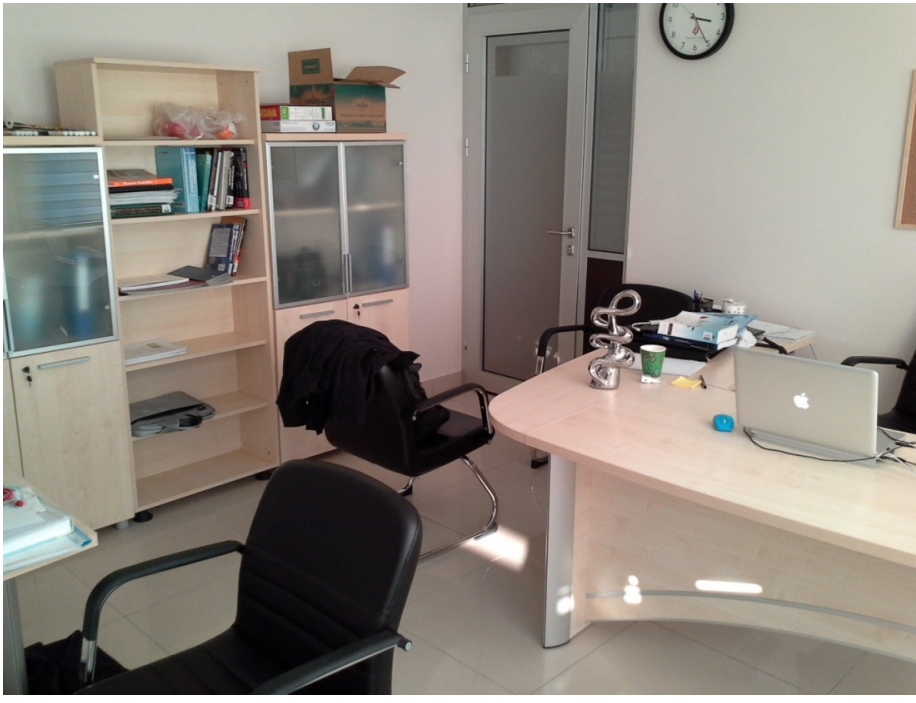

Fig. 9. The first frame of the Real Video Sequence.

[2] The details in [25] is not enough to implement the algorithm, yet we will attack the problem for comparison.

A Parameters

- SIFT peak threshold = 3
- SIFT edge threshold = 10
- SIFT feature elimination threshold = 5
- SIFT matching threshold = 2
- RANSAC iteration = 2,000
- Sampson error = 0.02
- Convolution kernel size = 60
- Convolution kernel, Gaussian standard deviation = 30
- Sliding window size = [64 48]
- Sliding window skip pixel = 5
- Specular classification weight = 0.001

References

1. Horn, B.: Shape from shading: A method for obtaining the shape of a smooth opaque object from one view (1970) Technical Report, MIT, Boston USA.
2. Horn, B.: Shape from shading information. (1975) New York USA: McGraw-Hill.
3. Koenderink, J., Doorn, A.V.: Photometric invariants related to solid shape. Optica Acta; **27** (1980) 981–996
4. Saint-Pierre, C., Boisvert, J., Grimard, G., Cheriet, F.: Detection and correction of specular reflections for automatic surgical tool segmentation in thoracoscopic images. Machine Vision and Applications **22** (2011) 171–180
5. Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., Fitzgibbon, A.: Kinectfusion : Real-time 3d 281 reconstruction and interaction using a moving depth. Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology ACM, New York, USA (2011) 559–568
6. Wolff, L., Boult, T.: Constraining object features using a polarization reflectance model. IEEE Trans. Pattern Analysis and Machine Intelligence **13** (1991) 635–657
7. Nayar, S., Fang, X., Boult, T.: Removal of specularities using color and polarization. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (1993) 583–590
8. Oren, M., Nayar, S.: A theory of specular surface geometry. Journal of Computer Vision **24** (1997) 105–124
9. Roth, S., Black, M.: Specular flow and the recovery of surface structure. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2006) 1869–1876
10. Nayar, S., Ikeuchi, K., Kanade, T.: Determining shape and reflectance of lambertian, specular, and hybrid surfaces using extended sources. International Workshop on Industrial Applications of Machine Intelligence and Vision (1989) 169–175
11. Adato, Y., Vasilyev, Y., Shahar, O.B., Zickler, T.: Toward a theory of shape from specular flow. International Conference on Computer Vision ICCV (2007) 1–8
12. Swaminathan, R., Kang, S., Szeliski, R., Criminisi, A., Nayar, S.: On the motion and appearance of specularities in image sequences. Lecture Notes Computer Science **2350** (2002) 508–523

13. Oo, T., Kawasaki, H., Ohsawa, Y., Ikeuchi, K.: The separation of reflected and transparent layers from real-world image sequence. Machine Vision and Applications **18** (2007) 17–24
14. Doerschner, K., Fleming, R., Yilmaz, O., Schrater, P., Kersten, D.: Visual motion and the perception of surface material. Current Biology (2011) 2010–2016
15. Yilmaz, O., Doerschner, K.: Detection and localization of specular surfaces using image motion cues. Machine Vision and Applications **25** (2014) 1333–1349
16. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60** (2004) 91–110
17. Velaldi, A., Fulkerson, B.: Vlfeat: An open and portable library of computer vision algorithms. Proceedings of the international conference on Multimedia, Firence, Italy (2010) 1469–1472
18. Beis, J., Lowe, D.: Shape indexing using approximate nearestneighbour search in high-dimensional spaces. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (1997) 1000–1006
19. Fischler, M., Bolles, R.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications ACM **24** (1981) 381–395
20. Hartley, R., Gupta, R., Chang, T.: Stereo from uncalibrated cameras. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (1992) 761–764
21. Sampson, P.: Fitting conic sections to very scattered data: An iterative refinement of the bookstein algorithm. Computational Graphics Image Processing **18** (1982) 97–108
22. Sharan, L., Rosenholtz, R., Adelson, E.H.: Material perception: What can you see in a brief glance? Journal of Vision **14** (2014)
23. Maenp, T.: The local binary pattern approach to texture analysis extensions and applications. Oulu Finland: Univeristy of Oulu (2003)
24. Heikkil, M., Pietikinen, M., Schmid, C.: Description of interest regions with local binary patterns. Pattern Recognition **42** (2009) 425–436
25. DelPozo, A., Savarese, S.: Detecting specular surfaces on natural images. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2007) 1–8