Extending Surf to the Color Domain

David Gossow, Peter Decker, and Dietrich Paulus, University of Koblenz-Landau (Germany)

Abstract

Automatic extraction of local features from images plays an important role in many computer vision tasks. During the last years, much focus has been put on making the features invariant to geometric transformations such as a rotation and scaling of the image. Recently, some work has been published concerning the integration of color information into the detection and description step of SIFT. In various evaluations, it has been shown that including color information can increase distinctiveness and invariance to photometric transformations caused by illumination changes. In this paper we build on the results from these approaches and apply them to the SURF descriptor, which is advantageous compared to SIFT in terms of speed, making it a perfect candidate for online applications, for example in the field of robotics. Our results show significant improvements concerning the repeatability and destinctiveness of SURF for 3d objects under varying illumination directions.

In contrast to many other evaluations we also determine the accuracy of the orientation assignment and include this into our comparisons.

INTRODUCTION

Although color cameras are widely spread nowadays, most popular state of the art feature detector and descriptor algorithms like SIFT [1] and SURF [2] still operate on intensity images only. It is obvious that by disregarding color values one loses information. An edge between a green and a blue patch for example would only provide the same information as between two shades of gray, other color edges might not be recognizable in a gray scale image at all. Additional color information could be useful in the detection step to identify salient interest points defined by a change in color. A descriptor encoding color information is expected to provide a higher distinctiveness than those based on intensity only. From these assumptions the following questions arise:

- How can color information be included in the detection to gain access to features whose localization is based on changes in color and not necessarily intensity?
- How can color information be included in the descriptor to improve the distinctiveness and hence recognizability of feature points?

In this paper we propose an extension to the SURF algorithm which works on color images and yields better results in terms of repeatability of the detector and distinctiveness of the descriptor.

The rest of the paper is organized as follows: The next section will give an overview of lately proposed methods for enhancing existing feature descriptors with color information. Following that, we discuss basic choices of color space, the inclusion of color in scale space representations as well as color invariants and color boosting. The next two chapters deal with the integration of color information into the detection step and descriptor respectively. They also contain an evaluation of our algorithm. The final chapter summarizes the paper.

Related Work

During the recent years there were several suggestions how to include color information into state of the art local interest point detectors and descriptors. Goedemé et.al. [3] propose an additional step after matching SIFT descriptors in which a 3 dimensional color descriptor based on global color moments is used to sort out wrong matches, i.e. matches for which the distance of the color descriptors exceeds a fixed threshold. Bosch et.al.[4] introduce a color SIFT descriptor using HSV color space, but for a dense sampling approach in contrast to sparse interest point detection. Burghouts and Geusebroek [5] compare several SIFT descriptor variations utilizing color features invariant to different photometric transformations. Van de Weijer et.al. show how to improve the saliency of detected interest points by using color information [6] and propose a color histogram based method to improve the SIFT descriptor [7]. Abdel-Hakim and Farag [8] propose an extension to the detection as well as the descriptor step of SIFT based on color invariants. An overview and evaluation of possible color invariants for descriptors is given by van de Sande et.al. in [9].

Using color images

In order to build a descriptor which is invariant to certain changes in illumination, we have to define the underlying reflectance model first. We assume Lambertian reflectance of surfaces and additional omnidirectional diffuse , which leads to an image creation process modeled as follows:

$$I_k(x,y) = \int e(\lambda)s(x,y,\lambda)\rho_k(\lambda)d\lambda + \int a(\lambda)\rho_k(\lambda)d\lambda \quad (1)$$

$e(\lambda)$ are the spectral characteristics of the light source, $s(x,y,\lambda)$ is the surface reflectivity on the point measured by the sensor at (x,y) and $\rho_k(\lambda)$ is the camera sensitivity curve for channel k. $a(\lambda)$ is the ambient term.

Possible color invariants

Many photometric invariants have been proposed to gain robustness or invariance towards changes in illumination [10, 11, 12, 13, 6]. We take a closer look at two invariants proposed in [10] which delivered the best results in recent evaluations [9, 5]. Both are defined on the scale space representation $L_k, k \in [1,2,3]$ of the image I_k:

$$L_k(\cdot,\sigma) = G(\cdot,\sigma) * I_k \quad ,$$

where G is a two-dimensional Gaussian kernel with variance σ, $*$ denotes the convolution and k the image channel.

W invariant: The W invariant is defined as the derivative of the image signal, normalized by the intensity channel L_1. It is invariant to local intensity changes, assuming planar surfaces which do not exhibit shading effects.

$$W_{k,x} = \frac{L_{k,x}}{L_1}$$

$$W_{k,xx} = \frac{L_{k,xx}}{L_1}$$

C invariant: The C invariant is defined as the derivative of the intensity-normalized color channel $\frac{L_k}{L_1}, k \in [2,3]$, thus being invariant to shadow and shading:

$$C_{k,x} = \frac{L_{k,x}L_1 - L_k L_{1,x}}{L_1^2}$$

$$C_{k,xx} = \frac{L_{k,xx}L_1^2 - L_k L_{1,xx}L_1 - 2L_{k,x}L_{1,x}L_1 + 2L_k L_{1,x}^2}{L_1^3}$$

$W_{k,y}$, $W_{k,yy}$, $W_{k,xy}$, $C_{k,y}$, $C_{k,yy}$ and $C_{k,xy}$ are constructed in analogy to the above definitions.

Since the invariants become unstable in dark regions of the image due to noise, we set the values as well as the derivatives to zero if the intensity is less than 5%.

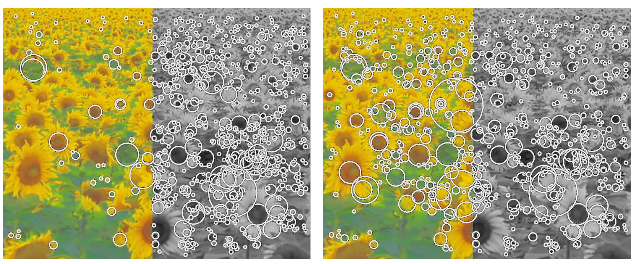
The choice of color space

We compared several color spaces which are related to RGB by a linear transform. All provide a channel that only contains intensity information, thus allow the calculation of the W and C invariants. *YCbCr* [14] is a native format for many cameras and thus directly accessible without the need of conversion. The *Gaussian color space* [15] is designed so its sensitivity curves approximate Gaussian functions. Thus, its color channels can be interpreted as Gaussian derivatives of the underlying light spectrum in the wavelength domain. Considering the derivative of normalized RGB [9] with the normalized channels $r = \frac{R}{R+G+B}$ and $g = \frac{G}{R+G+B}$ as a special case of the C invariant results in the *IRG color space*, where $I = R+G+B$. The *opponent color space* is introduced in [6]. The Gaussian, opponent and IRG color spaces are considered as they have shown superior performance compared to others like HSV in previous evaluations [5, 9].

Results of descriptor performance tests using these color spaces are part of the evaluation.

Color boosting

In [6], a method called color boosting is introduced, which aims at improving the distinctiveness of detected features by scaling the intensity and color channels of the opponent color space with respect to their information content. In [16], it is shown that combining a color-boosted detector with color-based SIFT features can achieve an improvement of up to 30% on classification compared to the intensity based SIFT. We propose a simplified alternative, scaling the intensity channel with a factor of 0.5 independent of the color space. Figure 1 illustrates the effect of this transformation on our interest point detector.



(a) without boosting (b) with boosting

Figure 1: Effect of color boosting on the detection of interest points.

Color in the detection step

Salient interest point detection in color scale space can be done by two different approaches: Using a single function that operates on all channels or treating the channels separately.

Ming et al. [17] proposed substituting the derivatives in the Hessian matrix by a weighted sum of the derivatives of all channels. However, Shi et al. [18] point out that this way, the derivatives may cancel out. Thus, the authors suggest the use of quaternions to represent the color information. However, the method relies on the calculation of the eigenvalues of the resulting quaternion Hessian matrix. This is computationally quite intensive [19] and thus not feasible for fast feature detection.

For our evaluation, we analyzed the channels separately for maxima of the determinant of the Hessian, just like the original SURF paper does on gray level images (called “colorSURF-separate” in the evaluation). If several maxima are found on the same location in scale space, the one with the highest value prevails and generates a feature. We compared this approach with examining the sum of determinants of the channels as a combining function (colorSURF-sum). This approach has the advantage that blobs with different polarities (e.g. light blobs on dark ground or dark blobs on light ground) in the different channels on the same spot will enforce each other instead of canceling out. The results are part of the evaluation of the descriptor repeatability.

Color feature descriptor

Sampling of color information

The intensity part of the descriptor is sampled from the intensity channel in 4×4 subregions, as described in [2]. For the sampling of the color information, we tried sampling from 1×1 up to 4×4 subregions. Since only the approach with just one subwindow was significantly worse than the rest, we chose to sample from 2×2 subregions. Figure 2 shows recall-precision curves for different numbers of subregions, tested on the ALOI data sets; the same evaluation on Mikolajczyk’s images gave similar results.

The color data of the 2 color channels is then appended to the descriptor formed from intensity data, thus yielding a vector of length 96.

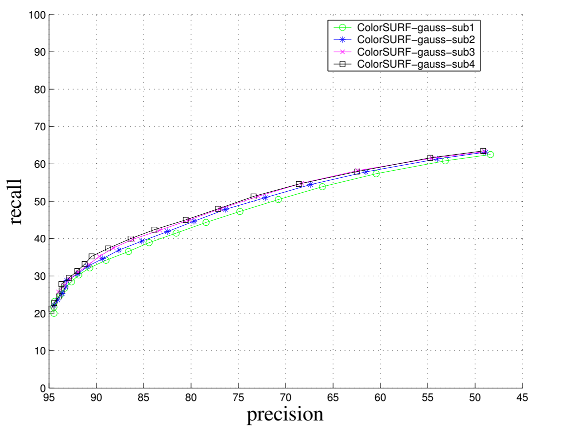


Figure 2: recall-precision for different numbers of subregions, evaluated on ALOI.

EXPERIMENTS AND EVALUATION

For our experiments, we used two different image databases provided by the community: The illumination direction collection of the Amsterdam Library of Object Images (ALOI [20]) and the evaluation framework provided by Mikolajczyk [21].

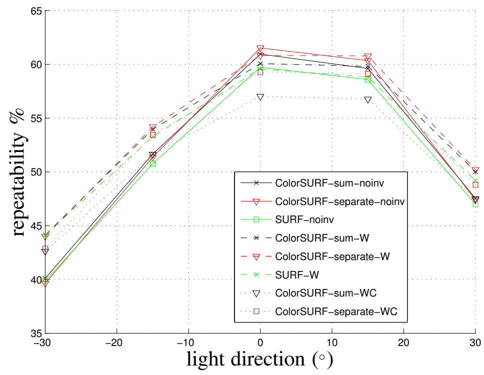


Figure 3: Repeatability of the detector under different illuminations using 100 objects from ALOI.

For our tests with ALOI, we picked a random set of 100 objects from the database, each containing six images. In theses image series the camera angle does not change, but the illumination azimuth is very different. Figure 11 shows an example. We use the image with all 5 lamps turned on (c1, l8) as reference and the others (c1, l1-l5) to measure performance.

From the framework of Mikolajczyk we used the graffiti (Figure 14) and bricks scene for tests under a change in viewpoint. The bark scene (Figure 12) has zoom and rotation. Since the other zoom/rotation scene from Mikolajczyk's data set is a gray scale image, we created another one containing an artificial affine transformation: the fields scene (Figure 13).

Unless otherwise stated, all tests have been performed using the Gaussian color space. The descriptor performance has been tested by matching the reference image with the second image of each image series (e.g. 15° light azimuth for the ALOI images).

Our implementation is based on a SURF implementation provided by Anael Orlinsky, which is part of the software Pan-o-matic [1]. We evaluated various open source implementations of SURF and found that it performs best, giving almost identical results to the binaries provided by the authors of [2].

Repeatability of the interest point detector

The most important aspect of an interest point detector is it's repeatability, which means that the same interest points in the world are detected again under different viewing angle and changes in illumination. We define repeatability as the quotient of recognized mutual interest points in the common region of two images and minimum number of features found in any of the images in that region. Interest points are considered as being recognized if their associated regions projected into one common image coordinate system overlap by a certain degree, as in [21].

We tested the repeatability on ALOI as well as Mikolajczyk's data to cover change in illumination, change in viewpoint as well as in plane rotation and zoom separately. The results suggest that color based detection treating the channels separately increases the stability of the detector in most cases. For large changes in illumination in the ALOI data, the detectors based on the W invariant outperform the non-invariant detectors especially for large illumination azimuths (see Figure 3). However, the repeatability drops on Mikolajczyk's data (Figure 4). The effect was most significant in presence of JPEG compression ("UBC" series) and underexposure ("Cars" series).

The color boosting in the detection step should yield more distinctive interest points when combined with a color descriptor. Thus, we test the performance of the $W_1C_2C_3$ descriptor for the

[1] http://aorlinsk2.free.fr/panomatic/

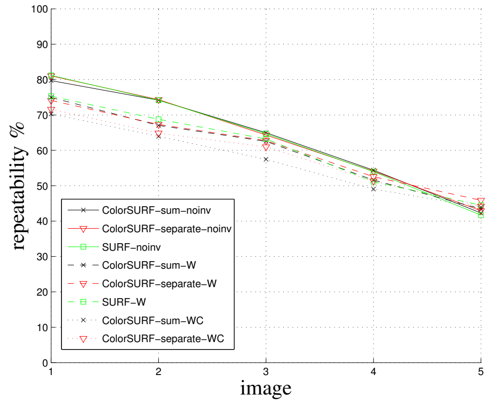
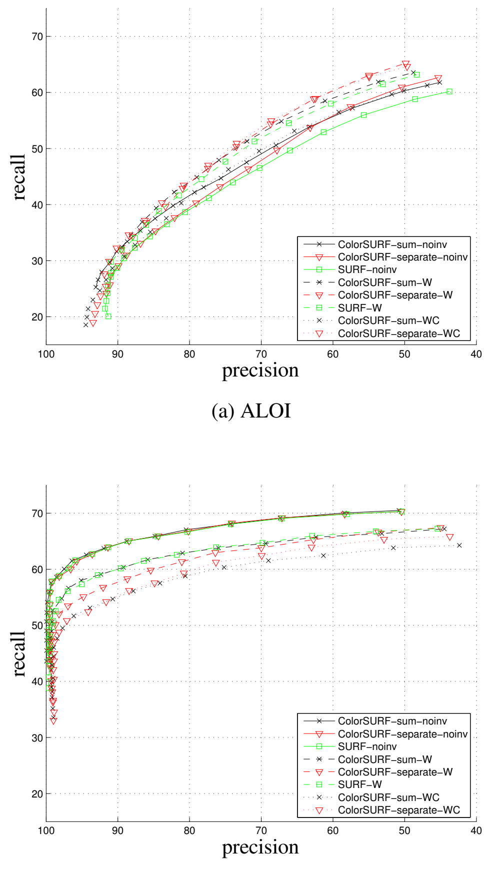


Figure 4: Mean Repeatability of the detector for Mikolajczyk's images



(a) ALOI

(b) Mikolajczyk

Figure 5: Performance of the $W_1C_2C_3$ descriptor when using different invariants in the detection phase.

interest points detected by the different detectors. Results show that this effect is very small (see Figure 5). When using the W invariant on Mikolajczyk's data, the inclusion of color information in the detector even decreases the performance. The W invariant with separate detection in all channels gives the best results on ALOI data, while the performance drops in Mikolajczyk's data with increasing invariance. Because the results are in general worse on the ALOI data, we favor the separate non-invariant approach for detection.

Accuracy of the orientation assignment

Assigning the correct orientation to a localized keypoint is crucial. An erroneous orientation directly influences the resulting descriptor and can lead to failures in matching. This has been neglected in most preceding analyzes of color extensions. To make sure that the orientation is still stable we analyzed the keypoints that were recognized again in the repeatability test for illumination change and checked for errors in orientation assignment. We compared different combinations of invariants (suffix -W and -WC in the graph) and two different approaches of combining the information from all channels. The first one builds one orientation histogram from all single channel gradients (suffix -separate) and the second one adds the gradients of all channels for each sample before building the histogram (suffix -sum).

Figure 6 shows that the summing of the W_1, C_2 and C_3 gradients yields slightly better results than the other methods on ALOI data, especially for large changes in illumination azimuth. Precision and recall of the descriptor based on the invariants W_1, C_2 and C_3 is improved accordingly on ALOI data, while for Mikolajczyk's images the performance slightly drops (see Figure 7).

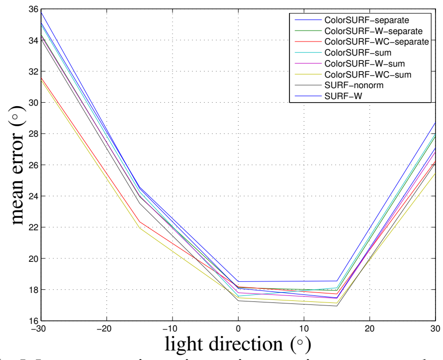


Figure 6: Mean error in orientation assignment under different illumination directions.

Descriptor performance

We compared the original gray level descriptor (entitled "SURF" in the graphs) to the color descriptor without additional invariance (ColorSURF) and the color descriptor based on the invariants W_1, W_2, W_3 (ColorSURF-W) and W_1, C_2, C_3 (ColorSURF-WC). The keypoints were detected using the original SURF algorithm.

We expect the WC-descriptor to improve in the ALOI images because of the invariance towards the present changes in illumination. Mikolajczyk's database does not contain changes in illumination geometry, thus adding invariance is not expected to improve results.

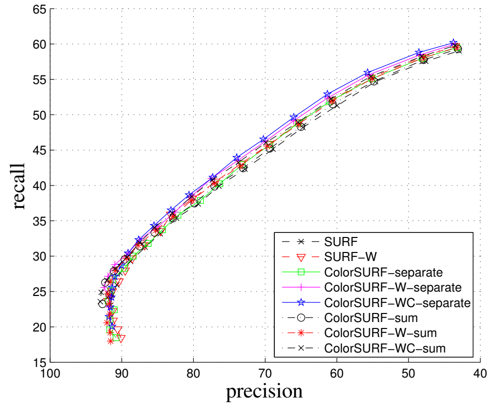
As Figure 8a shows, there is a notable improvement using the W and C invariants when dealing with 3d objects under changing illumination azimuth. The descriptor based solely on the W invariant yields a higher maximum precision than the combination of W_1 with C_2 and C_3. Adding invariance and color information has very little effect on the performance for the database of Mikolajczyk (see Figure 8b).

Performance of the different color spaces

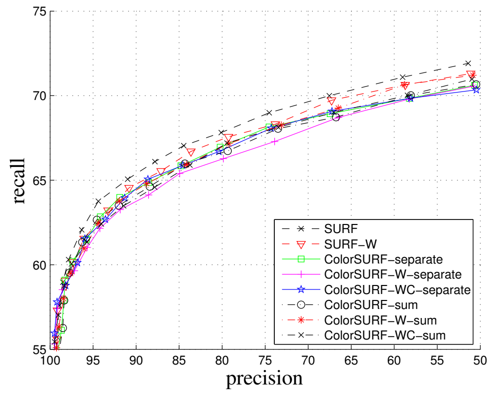
We computed recall-precision curves for the different color spaces on ALOI as well as Mikolajczyk's data sets. Results are shown in Figure 9 Except for IRG, all color spaces perform quite similar.

Final results

We compared our algorithm "colorSURF", using the $W_1W_2W_3$ invariants in detection and $W_1C_2C_3$ in orientation assignment and description to the original SURF implementation provided by the authors. Channels are treated separately in all stages. SURF-64 denotes the standard SURF descriptor and SURF-128 the extended descriptor of length 128. Our descriptor uses 2×2 subregions for the color channels, thus yielding a descriptor of length 96.



(a) ALOI



(b) Mikolajczyk

Figure 7: Descriptor Performance of the $W_1C_2C_3$ descriptor when using different methods of orientation assignment.

It can be seen that the performance for the different methods is very similar on Mikolajczyk's data, while under the illumination changes present in the ALOI data, the color based algorithm clearly outperforms the others (see figure 10).

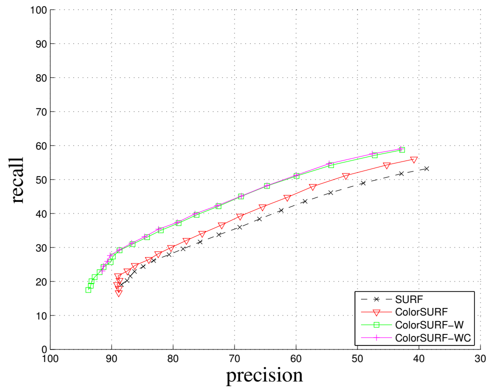
CONCLUSION

We proposed a method to integrate color information to the SURF detector and descriptor. We tested the influence of color space, color invariants and sampling size of the color descriptor separately, as well as different strategies for extrema detection in color scale space.

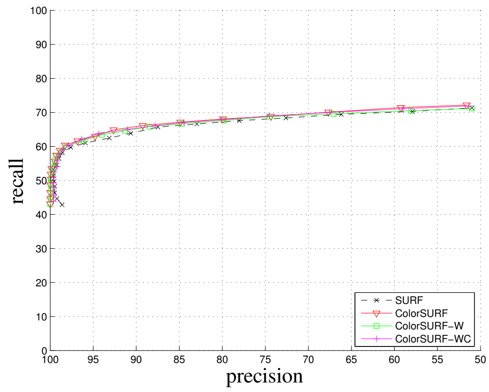
The evaluation revealed that integration of color information can improve the distinctiveness of the SURF descriptor. By using color features with certain invariances towards photometric effects, the robustness against changes in illumination azimuth is enhanced for 3d objects.

The combination of an upgraded detection and integration of color features results in an overall significant performance improvement for 3d objects under different lighting directions, while for constant lighting and geometric transforms the restrictions are minimal.

As color features impose only little overhead to the computation of SURF features, a color SURF computation can still be considered for time-critical vision tasks.

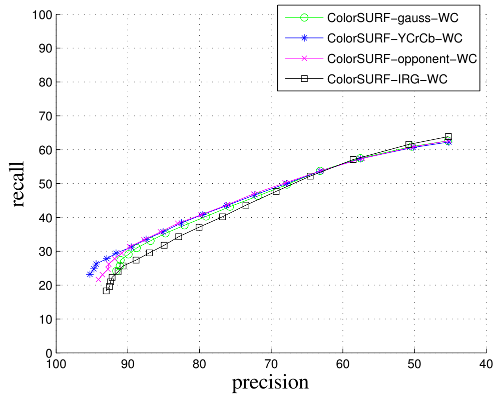


(a) ALOI

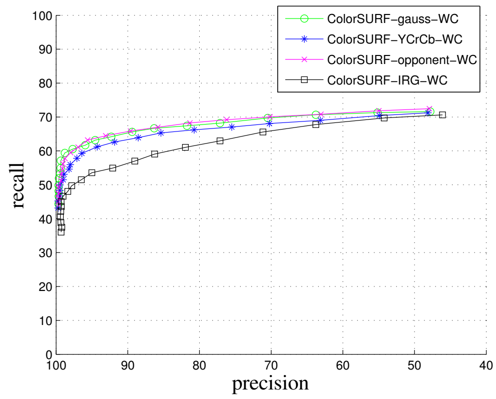


(b) Mikolajczyk

Figure 8: Descriptor performance using different invariants. On ALOI data, the invariant color descriptors show a significant improvement. On Mikolajczyk's data, the performance gain by using color information is very small.

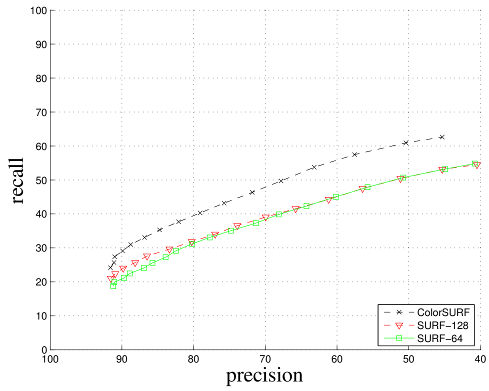


(a) ALOI

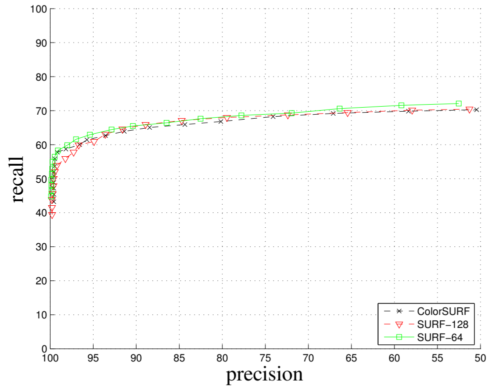


(b) Mikolajczyk

Figure 9: Descriptor performance for different color spaces.



(a) ALOI



(b) Mikolajczyk

Figure 10: Final comparison of our approach to the original SURF implementation. ColorSURF increases the performance of up to 25% compared to the original SURF when large changes in illumination azimuth are present, while performing similar otherwise.

Figure 11: Images from the ALOI [20] database used to evaluate the detector. For the first image (c1, 18), all five light sources are turned on. This is used as a reference. The others (c1, 11-15) are used to test the interest point detector and descriptor under a significant change in illumination azimuth.

Figure 12: Image series "Bark" from the evaluation framework provided by Mikolajczyk [21]. Image 1 is used as reference, the detector and descriptor performance under an in plane rotation and zoom is tested with the other images.

Figure 13: "Fields" image series containing artificial affine transformation.

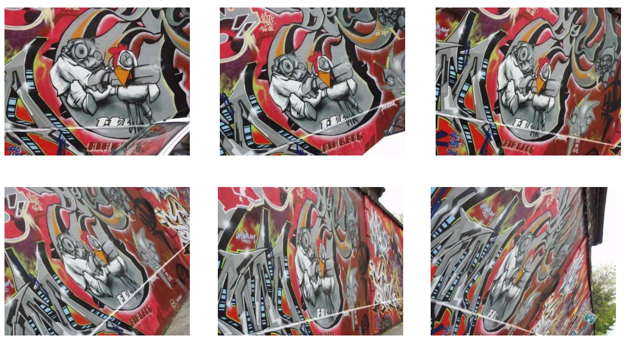


Figure 14: "Graffiti" image series from the evaluation framework provided by Mikolajczyk [21].

References

[1] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[2] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, "Surf: Speeded up robust features," *ECCV*, pp. 404–417, 2006.

[3] Toon Goedemé, Tinne Tuytelaars, and Luc Van Gool, "Omnidirectional sparse visual path following with occlusion-robust feature tracking," in *in: 6th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras, OMNIVIS05, in Conjunction with ICCV 2005*, 2005.

[4] Anna Bosch, Andrew Zisserman, and Xavier Munoz, "Scene classification using a hybrid generative/discriminative approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 4, pp. 712–727, 2008.

[5] Gertjan J. Burghouts and Jan-Mark Geusebroek, "Performance evaluation of local colour invariants," *Comput. Vis. Image Underst.*, vol. 113, no. 1, pp. 48–62, 2009.

[6] Joost van de Weijer, Theo Gevers, and Andrew D. Bagdanov, "Boosting color saliency in image feature detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 150–156, 2006.

[7] Joost van de Weijer and Cordelia Schmid, "Coloring local feature extraction," in *European Conference on Computer Vision*. 2006, vol. Part II, pp. 334–348, Springer.

[8] Alaa E. Abdel-Hakim and Aly A. Farag, "Csift: A sift descriptor with color invariant characteristics," *Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, vol. 02, pp. 1978–1983, 2006.

[9] Koen E. A. van de Sande, Theo Gevers, and Cees G. M. Snoek, "Evaluation of color descriptors for object and scene recognition.," in *CVPR*. 2008, IEEE Computer Society.

[10] Jan-Mark Geusebroek, Rein van den Boomgaard, Arnold W.M. Smeulders, and Hugo Geerts, "Color invariance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 12, pp. 1338–1350, 2001.

[11] Theo Gevers and Arnold W.M. Smeulders, "Color based object recognition," in *ICIAP '97: Proceedings of the 9th International Conference on Image Analysis and Processing-Volume I*, London, UK, 1997, pp. 319–326, Springer-Verlag.

[12] Todd Zickler, Satya P. Mallick, David J. Kriegman, and Peter N. Belhumeur, "Color subspaces as photometric invariants," *International Journal of Computer Vision*, vol. 79, no. 1, pp. 13–30, 2008.

[13] J. Van De Weijer, Th. Gevers, and J. M. Geusebroek, "Color edge detection by photometric quasi-invariants," in *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*, Washington, DC, USA, 2003, p. 1520, IEEE Computer Society.

[14] "Itu-r recommendation bt.709-5. parameter values for the hdtv standards for production and international programme exchange.," Tech. Rep., International Telecommunication Union, 2002.

[15] Jan-Mark Geusebroek, Rein van den Boomgaard, Arnold W.M. Smeulders, and Anuj Dev, "Color and scale: The spatial structure of color images," in *ECCV '00: Proceedings of the 6th European Conference on Computer Vision-Part I*, London, UK, 2000, pp. 331–341, Springer-Verlag.

[16] K. E. A. van de Sande, Th. Gevers, and C. G. M. Snoek, "Color descriptors for object category recognition," in *European Conference on Color in Graphics, Imaging and Vision*, 2008, pp. 378–381.

[17] Anlong Ming and Huadong Ma, "A blob detector in color images," in *CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval*, New York, NY, USA, 2007, pp. 364–370, ACM.

[18] Lilong Shi, Brian Funt, and Ghassan Hamarneh, "Quaternion color curvature," in *Sixteenth Color Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, 2008.

[19] Nicolas Le Bihan and S.J. Sangwine, "Quaternion principal component analysis of color images," 2003, pp. I: 809–812.

[20] J. M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders, "The amsterdam library of object images," *Int. J. Comput. Vis.*, vol. 61, no. 1, pp. 103–112, 2005.

[21] Krystian Mikolajczyk and Cordelia Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

Author Biography

David Gossow obtained a diploma in Computer Science from the University of Koblenz-Landau in 2009. He is now working there as a research officer in the active vision group. His primary research interests are robotics, color and computer vision.

Peter Decker obtained a diploma in Computer Science from the University of Koblenz-Landau in 2007. He is now working there as a PhD student in the active vision group. His major interests are structure from motion, camera pose estimation and 3d reconstruction from images.

Dietrich Paulus obtained a Bachelor degree in Computer Science from University of Western Ontario, London, Canada, followed by a diploma (Dipl.-Inf.) in Computer Science and a PhD (Dr.-Ing.) from Friedrich-Alexander University Erlangen-Nuremberg, Germany. He obtained his habilitation in Erlangen in 2001. Since 2001 he is at the institute for computational visualistics at the University Koblenz-Landau, Germany where he became a full professor in 2002. His primary interests are computer vision and robot vision.