

Maximum Detector Response Markers for SIFT and SURF

Florian Schweiger, Bernhard Zeisl, Pierre Georgel, Georg Schroth, Eckehard Steinbach, Nassir Navab

Technische Universität München

Email: {florian.schweiger, eckehard.steinbach, schroth}@tum.de
bernhard.zeisl@mytum.de
{georgel, navab}@in.tum.de

Abstract

In this paper, we introduce optimal markers to be used with the SIFT and SURF feature detectors. They can be applied to trigger the detection of feature points at desired locations. Unlike conventional marker systems, we do not propose a standalone solution comprising a set of markers and a thereto adapted detection algorithm. Instead, our markers are adapted to existing and established detectors. In particular, we introduce markers optimally suited for SIFT and SURF. We derive the optimal design and show their high detectability within a wide range of different imaging conditions in experiments on both synthetic and real data.

1 Introduction

In the field of Augmented Reality, the use of so called *fiducial markers* is common practice in order to obtain reliably extractable 2D-3D correspondences. Subsequent tasks such as pose estimation or object recognition are hence vastly simplified. Due to their versatility and low installation costs, fiducial markers are often chosen over tracking systems employing more costly infra-red or active markers. Example applications from other fields include visual servoing in robotic surgery and monitoring in production logistics.

In many applications, fiducial markers are supplemented with image features originating in salient points in the scene itself. Many algorithms exist for feature extraction. Maybe the most widely used are SIFT and SURF, two closely related algorithms that yield scale and rotation invariant image feature points. At the other end of the scale, in merely feature-based applications, it might even so be desirable to place some reliably detectable reference

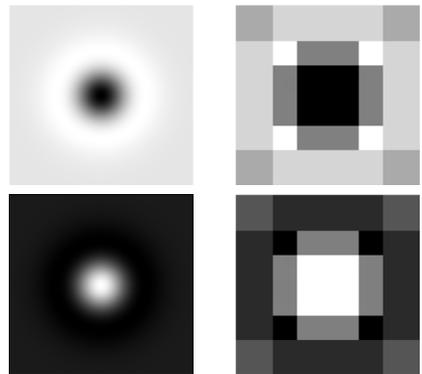


Figure 1: Our SIFT (left) and SURF markers.

points on unstructured surfaces, which would usually not be detected.

In this work, we propose a light-weight marker framework that conflates the 2-stage strategy consisting of marker detection and feature point extraction into a more efficient 1-step approach. We aim at the vast number of computer vision applications which are based on feature points, enhancing them with markers that fully integrate into the existing detection process. The markers we have developed for SIFT and SURF are provably optimal in that they trigger maximum response by the respective detector.

The remainder of the paper is structured as follows. In Section 2, we will give a brief overview on existing marker systems and the functioning of the SIFT and SURF feature detectors. Section 3 introduces the optimal markers for both these detectors, which are then validated experimentally in Section 4. We conclude the paper in Section 5 giving an outlook on future work.

2 Related Work

2.1 Fiducial Markers

Fiducial markers are widely used in a variety of different fields. Depending on the particular application, different designs have been proposed. An overview and comparison of different marker systems can be found in [13]. Fiducial marker systems typically consist of a set of distinguishable labels that are placed in the scene and can be detected and decoded by an associated algorithm. According to the design of the markers, the detection consists of several nontrivial steps such as edge detection, linking and line fitting. *Decoding* refers to the identification of a detected marker. Early systems used correlation-based approaches to match the appearance of a marker against a database of templates [4, 5]. State-of-the-art marker systems employ binary error correcting codes to allow the unique and robust identification of thousands of different markers [3]. Recently, there have been efforts towards lowering the computational complexity for mobile real-time applications [12, 11]. Compared to these highly specialized systems, our proposed marker scheme offers very elementary, yet highly valuable functionality: making feature points reliably detectable, at no additional expenses.

Desirable properties of a conventional marker system are low *false positive* and *false negative* rates, as well as low *inter-marker confusion* rates if the system comprises more than one distinguishable marker. That is, the system should neither report a detected marker when there is none, nor miss or mistake an actually present marker.

As mentioned before, state-of-the-art marker systems like, *e. g.*, *ARTag* [3] use error correction mechanisms which allow them to reach unrivaled performance in terms of false positive and inter-marker confusion rates. As our light-weight marker system provides only one marker, respectively two markers in the case of SURF, inter-marker confusion is not an issue. Our main goal is to have a highly detectable low-cost marker, not necessarily a uniquely detectable one.

The key virtue of our system is its remarkably low false negative rate. As we demonstrate in Section 4, it can be virtually guaranteed that our markers get detected even under dramatically varying imaging conditions. This makes them highly useful in applications where SIFT or SURF points need to

be found at a desired location.

2.2 Scale Invariant Feature Transform and Speeded Up Robust Features

Both SIFT [7] and SURF [1] belong to the family of *scale invariant feature detectors*. As such, they analyse an input image at different resolutions in order to repeatably find characteristic blob-like structures independently of their actual size in an image. To this end, both algorithms use multi-scale detection operators to analyze the so called *scale space representation* of an image. In a second step, detected features are assigned a rotation-invariant descriptor computed from the surrounding pixel neighborhood. The interested reader is referred to the original publications of SIFT and SURF for more detailed information about their inner workings. In the following, we will briefly review those parts of both algorithms which are relevant for our marker design: the detection operators, and the composition of the descriptor.

Detection Operators It has been shown that gradually filtering with a Gaussian kernel is particularly suited to build the scale space representation of natural images [6]. Furthermore, the Laplacian operator has proven very usable for interest point detection [8]. Consequently, SIFT uses an approximation to the Laplacian of Gaussians in order to find features on discrete scale levels. This is achieved by convolving the image I with *Difference-of-Gaussian* (DoG) filters of increasing size. The SIFT detection filter output reads:

$$D_{\text{SIFT}}(x, y, \sigma_k) = \left(H_{\text{DOG}}^k \star I \right) (x, y),$$

where $H_{\text{DOG}}^k = G^{\sigma_{k+1}} - G^{\sigma_k}$ (1)

Here, G^{σ_k} is a bivariate Gaussian with standard deviation σ_k representing the scale. Figure 2 shows the typical “flipped mexican hat” shape of H_{DOG}^k . The discrete scales are chosen appropriately to cover a reasonable range. Features are found as local maxima of the filter output in 3-dimensional scale space. Every feature point hence is a combination of location and scale maximizing D_{SIFT} .

SURF builds on the concepts of SIFT but introduces more radical approximations in order to speed up the detection process. Due to the use of integral images the complexity of SURF is greatly re-

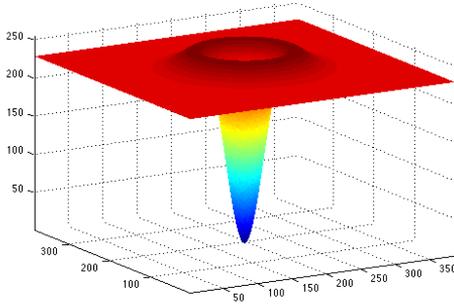


Figure 2: Difference of Gaussians (DoG) filter used by SIFT.

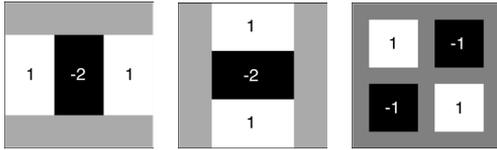


Figure 3: Box filteres used by SURF to approximate second order Gaussian derivatives.

duced, yet, SURF often achieves superior performance than its predecessor. Instead of the Laplacian operator, SURF uses the determinant of the Hessian for feature detection in scale space. Figure 3 shows the box filters \tilde{G}_{xx}^σ , \tilde{G}_{yy}^σ and \tilde{G}_{xy}^σ approximating the second order derivatives of the Gaussian G^σ . Scales are again discretized and depend on the size of the used box filters. By definition, a kernel size of $s \times s$ pixels corresponds to $\sigma = \frac{1.2}{9}s$. Applying these box filters to an image I yields the entries of the Hessian matrix, and we have the following detection operator output:

$$D_{\text{SURF}}(x, y, \sigma) = \det \begin{bmatrix} H_{11}(x, y) & H_{12}(x, y) \\ H_{21}(x, y) & H_{22}(x, y) \end{bmatrix} \quad (2)$$

where

$$H_{11} = \tilde{G}_{xx}^\sigma \star I, \quad H_{22} = \tilde{G}_{yy}^\sigma \star I, \\ H_{12} = H_{21} = \tilde{G}_{xy}^\sigma \star I$$

It is important to note here that, as opposed to the DoG filters, the SURF detection operator is nonlinear in the input image I .

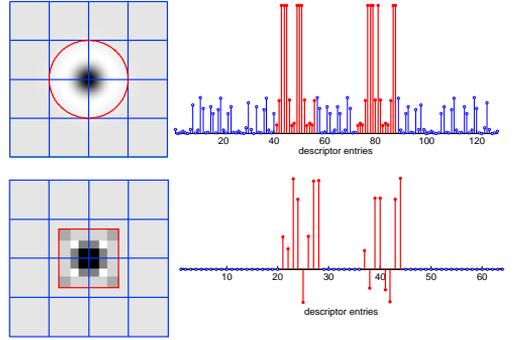


Figure 4: Left: The support used to compute the descriptors in size comparison with the respective markers. Right: The descriptors of the markers. The highlighted entries corresponding to the inner four subblocks constitute the marker's *signature*.

Descriptors After detection, every feature is assigned a "fingerprint" computed from the gradient distribution around its pixel location. The more distinct these *descriptors* are, the better the results obtained in a subsequent feature matching step. In our case, descriptors will also be used to identify markers in an image.

The SIFT descriptor is constructed from a square neighborhood of side length 12σ pixels, where σ is the scale of the feature. This neighborhood is aligned with the dominant local gradient direction, and Figure 4 shows its 4×4 subdivision. For each of the 16 subregions a 8-bin histogram of weighted gradients is built, and finally the descriptor is compiled by sorting the bins' contents from all subblocks into a vector of length $8 \times 4 \times 4 = 128$. Due to the scale and rotation adaptive creation process, SIFT descriptors are mostly invariant to moderate geometric transformations.

SURF uses a very similar descriptor layout, also based on a square region around the feature point which is aligned with the dominant gradient, and subdivided into 16 subblocks (see Fig. 4). The only differences are that the neighborhood is chosen 20σ pixels wide, and that every subblock only contributes 4 descriptor entries. Instead of histogram values, the sum and absolute sum of the gradient's x - and y -components are used. In total, the SURF descriptor hence comprises 64 entries.

3 Optimal Marker Design

In this section, we derive the optimal input images for the SIFT and SURF detectors, which then lend themselves as ideal markers for the respective detector. In this context, *optimal* refers to giving rise to the highest possible detector output.

3.1 The SIFT Marker

In case of SIFT, the task of determining the optimal marker is that of maximizing the output of a linear filter. As discussed in Section 2.2, this filter is a DoG as given in Equation (1).

In signal processing, *matched filters* are used to recover noisy signals of known shape [9]. A matched filter maximizes the signal-to-noise ratio, and its impulse response is simply the reversed version of the given signal. In our case, we are given a DoG filter and we search for the energy-limited signal which, superimposed with image noise, will yield maximum response at the filter output. In analogy to the matched filter design, we consequently choose our SIFT marker to be a DoG itself. Its shape and appearance are shown in Figures 2 and 1. Note that the reversed DoG is a valid “matched signal”, thus marker, as well.

Figure 4 shows the associated descriptor which can subsequently be used to identify the marker. Its characteristic shape due to the fact that the main lobes of the DoG fall inside the inner four subblocks is beneficial for descriptor matching. We will refer to these parts of the descriptor as the marker’s *signature*.

3.2 The SURF Marker

SURF on the other hand uses a nonlinear detection operator. Hence, the matched filter approach we took for SIFT is no longer applicable in this case. Nevertheless, we can derive the input image that will maximize the detector response by solving the corresponding optimization problem.

Without loss of generality, we will derive the SURF marker of size 9×9 pixels here which involves 81 variables in our optimization. The problem can of course also be solved for the other kernel sizes used by SURF, *i.e.*, 15, 21, 27, and so on. However, as the complexity of the problem increases with the square of the filter size, it should

be stated that the resulting markers will just be up-scaled versions of the 9×9 marker. Hence, we settle for the case $s = 9$ in the following, and consequently $\sigma = 1.2$. Assume we want (2) to attain a maximum at position (x_0, y_0) . Then the 81 pixels in the square region centered on (x_0, y_0) have to be taken into consideration. These are the only pixels which are actually covered by the support of the considered operator and their values must be adjusted such as to maximize $D_{\text{SURF}}(x_0, y_0, 1.2)$.

First, we arrange our 81 variables into the vector \mathbf{x} , and accordingly the entries of the box filters into the vectors \mathbf{g}_{xx} , \mathbf{g}_{xy} and \mathbf{g}_{yy} . From (2), the filter output can then be written as follows.

$$\begin{aligned} D_{\text{SURF}}(x_0, y_0, 1.2) &= \begin{vmatrix} \mathbf{g}_{xx}^\top \mathbf{x} & \mathbf{g}_{xy}^\top \mathbf{x} \\ \mathbf{g}_{xy}^\top \mathbf{x} & \mathbf{g}_{yy}^\top \mathbf{x} \end{vmatrix} \\ &= \mathbf{x}^\top \underbrace{(\mathbf{g}_{xx} \mathbf{g}_{yy}^\top - \mathbf{g}_{xy} \mathbf{g}_{xy}^\top)}_{\mathcal{G}} \mathbf{x} \\ &= \frac{1}{2} \mathbf{x}^\top (\mathcal{G}^\top + \mathcal{G}) \mathbf{x} \end{aligned}$$

With $\mathcal{A} = \mathcal{G}^\top + \mathcal{G}$, this leads to the following quadratic optimization problem if we additionally require that the pixel values be upper bounded.

$$\max_{\mathbf{x}} \mathbf{x}^\top \mathcal{A} \mathbf{x}, \quad \text{s.t. } \|\mathbf{x}\| \leq 1 \quad (3)$$

It can be shown that $\text{rank } \mathcal{A} = 3$, so the eigenvalue decomposition $\mathcal{A} = \mathcal{U} \Lambda \mathcal{U}^\top$ together with the substitution $\mathbf{y} = \mathcal{U}^\top \mathbf{x}$ yields the equivalent problem:

$$\begin{aligned} \max_{\mathbf{y}} \mathbf{y}^\top \Lambda \mathbf{y} &= \max_{\mathbf{y}} (\lambda_1 y_1^2 + \lambda_2 y_2^2 + \lambda_3 y_3^2), \\ \text{s.t. } \|\mathbf{y}\| &\leq 1 \end{aligned}$$

There are two solutions $\mathbf{y}_{\text{opt}} = [\pm 1, 0, \dots, 0]^\top$ and back-substitution reveals that \mathbf{x}_{opt} is the eigenvector corresponding to the largest eigenvalue of \mathcal{A} (and its inverse respectively). Rearranging \mathbf{x}_{opt} into a 9×9 image and adjusting its values to span the whole range of gray values eventually leads to the desired SURF markers, as depicted in Figures 5 and 1.

Due to their discrete nature, the SURF markers leave an even more characteristic *descriptor signature* than their SIFT counterparts (see Figure 4). It is particularly noteworthy that the signatures of the dark and light versions of the SURF marker are distinguishable. However, the entries corresponding to the sum of absolute gradients values are identical for both.

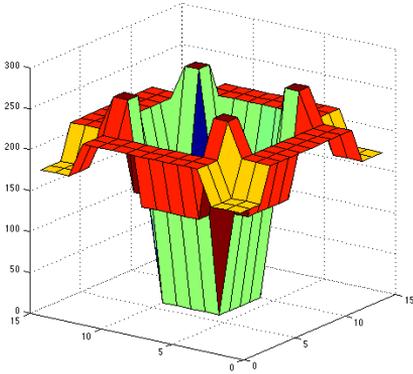


Figure 5: Maximum response SURF marker optimized for a kernel size of 15 pixels. Note the qualitative similarity to the reversed Mexican hat of Figure 2.

4 Experiments on Detectability

The markers we have derived in the previous section are provably optimal only if viewed under perfect conditions, *i. e.*, fronto-parallel to the camera, upright, and at the exact same size as the operator they were optimized for. The experiments we ran on synthetic data suggest, however, that our markers are still detectable if imaged at sizes halfway between scales, and under perspective distortions. In this section, we will give the results of our synthetic experiments, and demonstrate detectability also in real images. For all our experiments, we used the SIFT implementation by Andrea Vedaldi [10] and the OpenCV [2] implementation of SURF.

4.1 Detectability under Distortion

In general, there are two properties related to the performance of the proposed markers that have to be distinguished.

Detectability: A marker is *detectable* in an image if it generates a local maximum in scale space, *i. e.*, the detector will report a feature point at the marker position. This is the minimum requirement towards our markers.

Unique detectability: If the imaged marker even triggers the global maximum in scale space, or if the combination of high detector re-



Figure 6: Examples of the distortions we applied to the SURF marker in our synthetic experiments. In plane rotation by 10° , out-of-plane rotation by 40° , Gaussian noise with standard deviation equal to 50 gray values. The green cross indicates the position at which the marker has been detected

sponse and signature similarity identifies it as a marker, it is *uniquely detectable*. The experiments we describe in Section 4.2 suggest that our markers also have this property.

The goal of the following experiments is to backup the theoretical optimality of our marker design. More specifically, we want to show that even under unfavorable viewing conditions, *i. e.*, other than those assumed during the derivations, the markers still yield over-average detector response. What we would ideally expect is a detector response invariant to all these effects.

We used the following setup: A synthetically generated marker, either SIFT or SURF, is placed centered in front of a uniform background in an image of size 513×513 pixels. The initial size of the SURF marker is $s = 147$ pixels, and $\sigma = 15$ pixels in the SIFT case. After applying different distortions to the image, we run the respective feature detector and measure the detector response. At the same time, we follow the localization error with respect to ground truth, and if the marker is detected more than 3 pixels off, we declare it undetected. This is a rather strict rule compared to the used marker size. The distortions we have investigated are

- (a) *scaling*,
- (b) *in-plane rotation*,
- (c) *out-of plane rotation* and
- (d) *image noise*.

Figure 6 illustrates the applied distortions in the case of the SURF marker. Note that for the depicted out-of-plane rotation, the detection of the marker did not pass our 3 pixel accuracy test. Yet, the detection itself was apparently correct.

Figures 7 to 9 show the behavior of the response values for selected examples. Our overall observa-

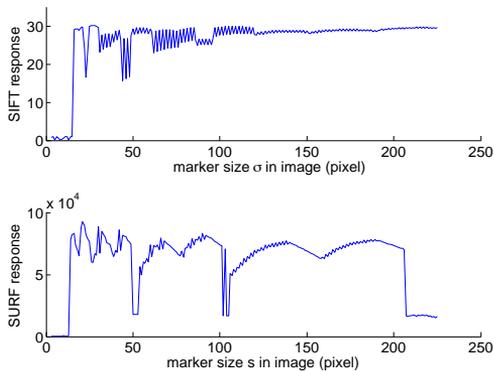


Figure 7: Behavior of detector response under scaling of the marker.

tion is that the detector response in general stays on a fairly high level, *i. e.*, close to the maximum value reachable in the absence of distortion. Within the plotted ranges the 3 pixel threshold was never violated (except for major angles in the out-of-plane rotation scenario). However, there are some unexpected effects which require further investigations.

A first interesting observation is the behavior of the response over the actual marker scale (Fig. 7). Both SURF and SIFT exhibit isolated ditches at mid-scales. Apparently, the spacing between neighboring scale values is adversely coarse in these regions. Apart from that, the two curves show the expected behavior. At too small sizes, inferior to the smallest detector operator, the marker does not get detected at all. In the case of SURF, there is also an upper limit to the detectable marker size as, different from SIFT, there is a predefined number of scale steps. This experiment suggests a minimum size for the SURF marker of 12 pixels, and a maximum size of 210 pixels. In practice, typical sizes are likely not to exceed 50 pixels as markers are preferred to be as nonintrusive as possible.

Regarding the behavior with respect to in-plane rotation, the results are convincing (Fig. 8). Even though there are minor variations in the detector output, the overall response level remains constantly high. This was to be expected for the rotationally symmetric SIFT marker, because the used resolution is sufficiently high to avoid severe aliasing artifacts. It is remarkable in case of the SURF

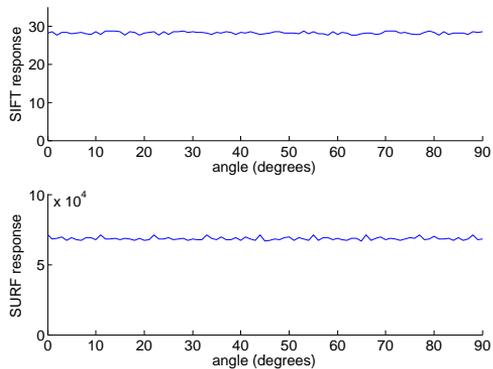


Figure 8: Behavior of detector response under in-plane rotation.

marker however, for which reduced detectability for rotation angles around 45° would seem inevitable.

For out-of-plane rotations (Fig. 9), the markers hit the limits that are given by the feature detectors themselves. It is known that both detectors do not cope well with angles beyond some 40 degrees. While the SIFT marker does well in terms of high detector output, it exceeds the 3px localization accuracy test for angles greater than 35 degrees. The SURF marker exhibits a significant response decrease and also deviates from ground truth more than 3 pixels for angles superior to 25 degrees.

The behavior with respect to artificially added image noise is not shown here. It turns out that both marker variants are highly robust against noise. The performance degradation follows a linear descent and is only relevant for unnaturally high noise levels with a standard deviation of at least 20% of the range of gray values.

4.2 Example Applications

Our markers are particularly useful whenever homogeneous, unstructured surfaces need to be "spiced up" with detectable features. Here, we touch on two example applications .

Robotic Navigation Imagine an experimental setup where a solely vision-based robot is to navigate through the lab in order to fulfill a certain task. Without a map or detectable markers in the

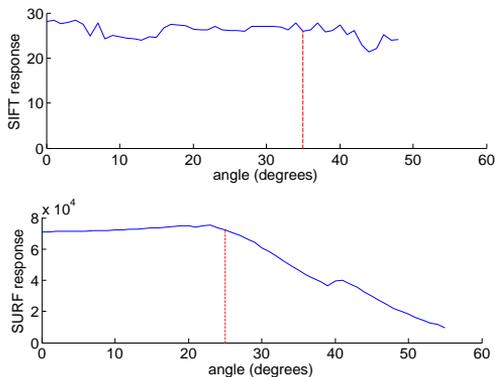


Figure 9: Behavior of detector response under perspective distortion. The dotted red lines indicate the limit above which the localization error exceeds 3 px.

scene, the robot would certainly lose track. Assume the task involves the recognition of certain objects which is based on the use of SURF features already. Such a prototypical scenario is the ideal application for our markers: We lay out the path to be followed by simply lining it with markers the robot can detect with its already built in functionalities.

Given the sensitivity to out-of-plane rotations according to our experiments from Section 4.1, we propose a simple preprocessing step for this particular scenario: As the height of the robot mounted camera remains constant, we can assume that the markers are always perceived perspectively distorted in the same way, *i. e.*, they appear wider than high. We found that stretching the image vertically by a factor 2 (using bilinear interpolation) improves marker detection significantly.

As illustrated in Figure 10, we placed 14 8×8 cm markers on the floor and ran SURF with very low response threshold on the depicted image. In total, some 30000 features were detected. Nevertheless, the markers derived in this paper were reliably found as those features with the highest detector response values. Two of the markers were detected twice, each time with different orientation, so there are 16 highest response features for 14 present markers.

Figure 11 shows the response values in descending order. The labels in Figure 10 relate to this ranking. A drop from feature 16 to feature 17 is clearly



Figure 10: Image captured with a Panasonic DMC-FX1 at resolution 2048×1536 pixels, after vertical upsampling by factor 2. The detected SURF markers are labelled according to the ranking in Figure 11.

visible., *i. e.*, the markers are indeed *uniquely* detectable in the image. Figure 12 shows the similarity between the descriptor signature of detected points and the signature of the dark marker template. Again, the markers clearly stand out. Note that there are other features, *e. g.*, the 56th, with similar signatures, but their rank, and equivalently the detector response they triggered are much lower.

A property of this example setup that is worthwhile mentioning is that since every detected marker comes with a scale assigned by the SURF algorithm, the robot can already make first assumptions about the marker distances without the need for stereo vision and triangulation.

SIFT/SURF CAVE Another example application that takes a similar line is to have a wallpaper or poster textured with the markers from Section 3. By

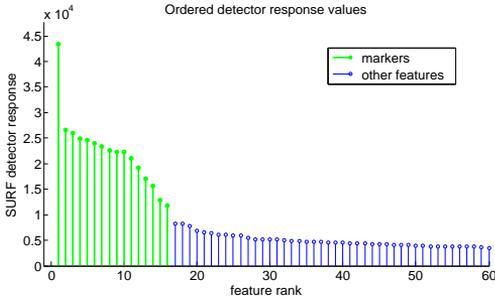


Figure 11: The SURF detector responses from the example in Figure 10 in descending order. The first 16 values belong to the imaged markers.

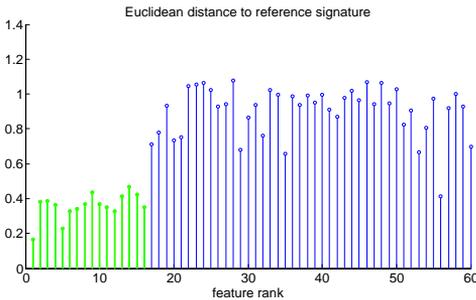


Figure 12: Signature similarity for the SURF features from the example in Figure 10. For every feature, the Euclidean distance between its signature and the signature of the template marker is plotted.

simply putting up such posters, a computer vision lab can easily be turned into a “SIFT/SURF CAVE” with dense feature points all over the walls.

Mixing light and dark versions of the marker, such posters can carry one bit per marker allowing, *e. g.*, to encode information about different parts of the room. In principle, there are two ways to distinguish both marker variants from each other, either based on marker signature or by means of the Laplacian of the respective marker, *i. e.*, the sum of second derivatives. For SIFT, it is precisely the detector output (1) that approximates the Laplacian. In the case of SURF, it is given by the trace of the Hessian matrix introduced in (2). The Laplacian is strictly negative for dark markers and strictly positive for light ones, hence a simple thresholding is

sufficient for separation. We found that this approach is more reliable than signature comparison, and we will use it in the following.

We illustrate the combination of markers using arrays composed of three SURF markers each. Figure 13 shows two such super-markers printed out on A4 paper and attached in our lab. The upper left corner of the two-by-two arrays is left blank to have an obvious array orientation. This design allows eight different marker arrays which could be used to tag strategic points in the room.

We propose the following simple algorithm to identify and decode these marker arrays. First, the SURF markers are uniquely detected in the image by means of their outstanding detector response and signature similarity. The feasibility of this step will be demonstrated shortly. Second, those markers are identified whose closest two neighbors satisfy certain geometric constraints, namely (a) their respective distances are similar and (b) they lie in roughly perpendicular directions. This yields 3-marker clusters in linear time with regard to the number of markers in place. Assuming that all the marker arrays are imaged close to upright, the topmost marker of each triple is then defined to carry the first bit, the leftmost one the second and the remaining marker the third bit. Finally, the bit value of each marker is determined using the sign of its Laplacian (see Figure 15). This approach is supposed to illustrate the combination of markers and can clearly be extended in a more sophisticated way.

Figure 14 shows the results of the SURF detection step. Similarly to the robot navigation example, we examine the found features in terms of detector response and distance to the reference signature from Figure 4. Every point in the figure corresponds to a detected feature, and obviously, the six features that belong to the markers form a cluster in an area with high detector response and low signature distance. As opposed to the previous experiment where only dark markers were deployed, we aim at simultaneously identifying both versions of the SURF marker in this example. Therefore, we decided not to work with the full marker signatures, but only those entries which are common to both marker variants. In the OpenCV implementation of SURF the corresponding descriptor dimensions are 23, 24, 27, 28, 39, 40, 43 and 44. We call this reduced representation *version independent signature (VIS)*.



Figure 13: Two out of eight possible 3-bit marker arrays photographed with a Sony DSC-S85 at resolution 2272×1704 pixels. The location of the six SURF features which yielded highest detector response are displayed as green crosses.

The distribution of response-distance pairs in Figure 14 suggests the use of a more complex classification algorithm than the current one which is primarily based on detector response. Instead, an oblique separation line seems more suited to the problem. In future work, it should be investigated to what extent other linear or nonlinear classification approaches can improve marker identification.

5 Conclusion

We have presented our work on maximum detector response markers for SIFT and SURF. We derived their optimal design theoretically, and presented first experimental results. The markers we propose are inexpensive and easy to use. It is sufficient to print them out and arbitrarily place them in the scene. No additional detection algorithms on top of SIFT or SURF are required. The stated algorithms detect our markers under a wide range of viewpoints, such that they can be uniquely identified due to their high response and their discriminative descriptor signature.

We plan to conduct more exhaustive experiments on real data in order to assess the markers' performance quantitatively. Especially a sound evaluation of the localization error in natural applications will be addressed in future work.

If you are interested in testing the markers we

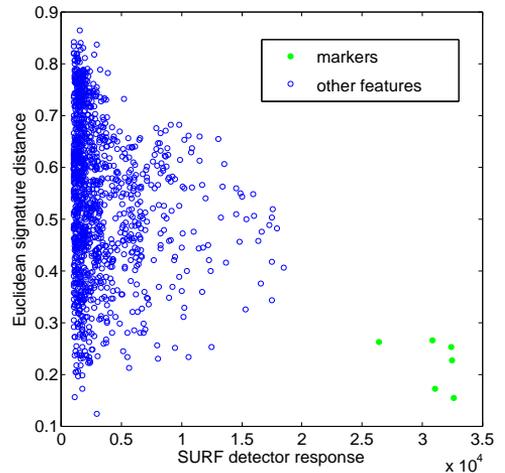


Figure 14: Detector response and VIS distance values for all the SURF features extracted from the image in Figure 13. The markers form an apparent cluster.

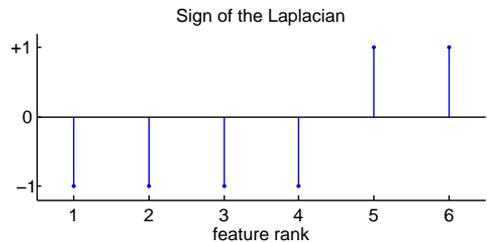


Figure 15: The sign of the Laplacian allows to reliably distinguish the light (+1) and dark (-1) versions of the SURF marker (compare with Figure 13).

proposed in this paper yourself, we invite you to visit our homepage¹ and download the markers in PDF format. Make sure you print them in high quality as this is essential for optimal results! The markers we used in the real world experiments from Section 4.2 were printed on a HP Color LaserJet CP4005 at 1200 dpi using bright-white paper. Please feel free to share your experiences with us!

¹<http://www.lmt.ei.tum.de/florian/markers>

References

- [1] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. *Lecture Notes in Computer Science*, 3951:404, 2006.
- [2] G. Bradski and A. Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media, Inc., 2008.
- [3] M. Fiala. ARTag, a fiducial marker system using digital techniques. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005*, volume 2, 2005.
- [4] H. Kato and M. Billinghurst. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In *2nd IEEE and ACM International Workshop on Augmented Reality, 1999.(IWAR'99) Proceedings*, pages 85–94, 1999.
- [5] H. Kato, M. Billinghurst, I. Poupyrev, K. Imamoto, and K. Tachibana. Virtual object manipulation on a table-top AR environment. In *IEEE and ACM International Symposium on Augmented Reality, 2000.(ISAR 2000). Proceedings*, pages 111–119, 2000.
- [6] T. Lindeberg. Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of applied statistics*, 21(1):225–270, 1994.
- [7] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [8] K. Mikolajczyk. Detection of local features invariant to affine transformations. *PhD thesis, Institut National Polytechnique de Grenoble, France*, 2002.
- [9] G. Turin. An introduction to matched filters. *Information Theory, IRE Transactions on*, 6(3):311–329, 1960.
- [10] A. Vedaldi. An open implementation of the SIFT detector and descriptor. Technical report, 070012, UCLA CSD, 2007.2.
- [11] D. Wagner, T. Langlotz, and D. Schmalstieg. Robust and Unobtrusive Marker Tracking on Mobile Phones. In *7th IEEE/ACM International Symposium on Mixed and Augmented Reality, 2008. ISMAR 2008*, pages 121–124, 2008.
- [12] D. Wagner and D. Schmalstieg. Artoolkit-plus for pose tracking on mobile devices. In *Proceedings of 12th Computer Vision Winter Workshop (CVWW'07)*, pages 139–146, 2007.
- [13] X. Zhang, S. Fronz, and N. Navab. Visual marker detection and decoding in ar systems: A comparative study. In *Proceedings of the 1st International Symposium on Mixed and Augmented Reality*. IEEE Computer Society Washington, DC, USA, 2002.