Herbert Bay[1], Tinne Tuytelaars[2], and Luc Van Gool[1,2]

[1] ETH Zurich
{bay, vangool}@vision.ee.ethz.ch
[2] Katholieke Universiteit Leuven
{Tinne.Tuytelaars, Luc.Vangool}@esat.kuleuven.be

Abstract. In this paper, we present a novel scale- and rotation-invaria interest point detector and descriptor, coined SURF (Speeded Up R bust Features). It approximates or even outperforms previously propos schemes with respect to repeatability, distinctiveness, and robustness, can be computed and compared much faster.

This is achieved by relying on integral images for image convolutio by building on the strengths of the leading existing detectors and descr tors (*in casu*, using a Hessian matrix-based measure for the detector, a a distribution-based descriptor); and by simplifying these methods to t essential. This leads to a combination of novel detection, description, a matching steps. The paper presents experimental results on a standa evaluation set, as well as on imagery obtained in the context of a real- object recognition application. Both show SURF's strong performanc

1 Introduction

The task of finding correspondences between two images of the same object is part of many computer vision applications. Camera calibra reconstruction, image registration, and object recognition are just a search for discrete image correspondences – the goal of this work – c vided into three main steps. First, 'interest points' are selected at d locations in the image, such as corners, blobs, and T-junctions. The m able property of an interest point *detector* is its repeatability, i.e. w reliably finds the same interest points under different viewing conditio the neighbourhood of every interest point is represented by a feature ve *descriptor* has to be distinctive and, at the same time, robust to nois tion errors, and geometric and photometric deformations. Finally, the d vectors are *matched* between different images. The matching is often b distance between the vectors, e.g. the Mahanalobis or Euclidean dista dimension of the descriptor has a direct impact on the time this tak lower number of dimensions is therefore desirable.

It has been our goal to develop both a detector and descriptor, comparison to the state-of-the-art are faster to compute, while not s performance. In order to succeed, one has to strike a balance between t

A. Leonardis, H. Bischof, and A. Pinz (Eds.): ECCV 2006, Part I, LNCS 3951, pp. 404–

the literature (e.g. [1, 2, 3, 4, 5, 6]). Also, detailed comparisons and evalu
benchmarking datasets have been performed [7, 8, 9]. While constructin
detector and descriptor, we built on the insights gained from this prev
in order to get a feel for what are the aspects contributing to perforn
our experiments on benchmark image sets as well as on a real object re
application, the resulting detector and descriptor are not only faster,
more distinctive and equally repeatable.

When working with local features, a first issue that needs to be
the required level of invariance. Clearly, this depends on the expectec
ric and photometric deformations, which in turn are determined by th
changes in viewing conditions. Here, we focus on scale and image rotat
ant detectors and descriptors. These seem to offer a good compromise
feature complexity and robustness to commonly occurring deformatio
anisotropic scaling, and perspective effects are assumed to be second
fects, that are covered to some degree by the overall robustness of the d
As also claimed by Lowe [2], the additional complexity of full affine-inva
tures often has a negative impact on their robustness and does not pay
really large viewpoint changes are to be expected. In some cases, even
invariance can be left out, resulting in a scale-invariant only version
scriptor, which we refer to as 'upright SURF' (U-SURF). Indeed, in q
applications, like mobile robot navigation or visual tourist guiding, th
often only rotates about the vertical axis. The benefit of avoiding the
rotation invariance in such cases is not only increased speed, but also
discriminative power. Concerning the photometric deformations, we
simple linear model with a scale factor and offset. Notice that our det
descriptor don't use colour.

The paper is organised as follows. Section 2 describes related work,
our results are founded. Section 3 describes the interest point detectio
In section 4, the new descriptor is presented. Finally, section 5 shows t
imental results and section 6 concludes the paper.

2 Related Work

Interest Point Detectors. The most widely used detector probably is
ris corner detector [10], proposed back in 1988, based on the eigenval
second-moment matrix. However, Harris corners are not scale-invari
deberg introduced the concept of automatic scale selection [1]. This
detect interest points in an image, each with their own characteris
He experimented with both the determinant of the Hessian matrix
the Laplacian (which corresponds to the trace of the Hessian matrix)
blob-like structures. Mikolajczyk and Schmid refined this method, cr
bust and scale-invariant feature detectors with high repeatability, w

mated the Laplacian of Gaussian (LoG) by a Difference of Gaussian
filter.

Several other scale-invariant interest point detectors have been prop
amples are the salient region detector proposed by Kadir and Brady [1
maximises the entropy within the region, and the edge-based region dete
posed by Jurie *et al.* [14]. They seem less amenable to acceleration thou
several affine-invariant feature detectors have been proposed that can c
longer viewpoint changes. However, these fall outside the scope of this

By studying the existing detectors and from published comparison
we can conclude that (1) Hessian-based detectors are more stable an
able than their Harris-based counterparts. Using the determinant of th
matrix rather than its trace (the Laplacian) seems advantageous, as it
on elongated, ill-localised structures. Also, (2) approximations like the
bring speed at a low cost in terms of lost accuracy.

Feature Descriptors. An even larger variety of feature descriptors
proposed, like Gaussian derivatives [16], moment invariants [17], com
tures [18, 19], steerable filters [20], phase-based local features [21], and
tors representing the distribution of smaller-scale features within th
point neighbourhood. The latter, introduced by Lowe [2], have been
outperform the others [7]. This can be explained by the fact that they
a substantial amount of information about the spatial intensity patter
at the same time being robust to small deformations or localisation er
descriptor in [2], called SIFT for short, computes a histogram of local
gradients around the interest point and stores the bins in a 128-dim
vector (8 orientation bins for each of the 4×4 location bins).

Various refinements on this basic scheme have been proposed. Ke
thankar [4] applied PCA on the gradient image. This PCA-SIFT yie
dimensional descriptor which is fast for matching, but proved to be les
tive than SIFT in a second comparative study by Mikolajczyk *et al.* [8] a
feature computation reduces the effect of fast matching. In the same
the authors have proposed a variant of SIFT, called GLOH, which pro
even more distinctive with the same number of dimensions. However,
computationally more expensive.

The SIFT descriptor still seems to be the most appealing descriptor
tical uses, and hence also the most widely used nowadays. It is distin
relatively fast, which is crucial for on-line applications. Recently, Se
implemented SIFT on a Field Programmable Gate Array (FPGA) and
its speed by an order of magnitude. However, the high dimensionality
scriptor is a drawback of SIFT at the matching step. For on-line ap
on a regular PC, each one of the three steps (detection, description, m
should be faster still. Lowe proposed a best-bin-first alternative [2] in
speed up the matching step, but this results in lower accuracy.

very basic Laplacian-based detector. It relies on integral images to re
computation time and we therefore call it the 'Fast-Hessian' detector
scriptor, on the other hand, describes a distribution of Haar-wavelet
within the interest point neighbourhood. Again, we exploit integral i
speed. Moreover, only 64 dimensions are used, reducing the time for fea
putation and matching, and increasing simultaneously the robustness
present a new indexing step based on the sign of the Laplacian, which
not only the matching speed, but also the robustness of the descriptor

In order to make the paper more self-contained, we succinctly discus
cept of integral images, as defined by [23]. They allow for the fast implen
of box type convolution filters. The entry of an integral image $I_{\Sigma}(\mathbf{x})$ at a
$\mathbf{x} = (x, y)$ represents the sum of all pixels in the input image I of a re
region formed by the point $\mathbf{x}$ and the origin, $I_{\Sigma}(\mathbf{x}) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i,$
I_{Σ} calculated, it only takes four additions to calculate the sum of the i
over any upright, rectangular area, independent of its size.

3 Fast-Hessian Detector

We base our detector on the Hessian matrix because of its good perfor
computation time and accuracy. However, rather than using a different
for selecting the location and the scale (as was done in the Hessia
detector [11]), we rely on the determinant of the Hessian for both. Give
$\mathbf{x} = (x, y)$ in an image I, the Hessian matrix $\mathcal{H}(\mathbf{x}, \sigma)$ in $\mathbf{x}$ at scale σ
as follows

$$\mathcal{H}(\mathbf{x}, \sigma) = \begin{bmatrix} L_{xx}(\mathbf{x}, \sigma) & L_{xy}(\mathbf{x}, \sigma) \\ L_{xy}(\mathbf{x}, \sigma) & L_{yy}(\mathbf{x}, \sigma) \end{bmatrix},$$

where $L_{xx}(\mathbf{x}, \sigma)$ is the convolution of the Gaussian second order
$\frac{\partial^2}{\partial x^2} g(\sigma)$ with the image I in point $\mathbf{x}$, and similarly for $L_{xy}(\mathbf{x}, \sigma)$ and L

Gaussians are optimal for scale-space analysis, as shown in [24]. In
however, the Gaussian needs to be discretised and cropped (Fig. 1 left
even with Gaussian filters aliasing still occurs as soon as the resulting i
sub-sampled. Also, the property that no new structures can appear whil
lower resolutions may have been proven in the 1D case, but is known to
in the relevant 2D case [25]. Hence, the importance of the Gaussian seem
been somewhat overrated in this regard, and here we test a simpler al
As Gaussian filters are non-ideal in any case, and given Lowe's success
approximations, we push the approximation even further with box filte
right half). These approximate second order Gaussian derivatives, an
evaluated very fast using integral images, independently of size. As sho
results section, the performance is comparable to the one using the d
and cropped Gaussians.

Fig. 1. Left to right: The (discretised and cropped) Gaussian second ord
derivatives in y-direction and xy-direction, and our approximations thereof
filters. The grey regions are equal to zero.

The 9×9 box filters in Fig. 1 are approximations for Gaussian seco
derivatives with $\sigma = 1.2$ and represent our lowest scale (i.e. highes
resolution). We denote our approximations by D_{xx}, D_{yy}, and D_{xy}. Th
applied to the rectangular regions are kept simple for computational
but we need to further balance the relative weights in the expressio
Hessian's determinant with $\frac{|L_{xy}(1.2)|_F |D_{xx}(9)|_F}{|L_{xx}(1.2)|_F |D_{xy}(9)|_F} = 0.912... \simeq 0.9$, whe
the Frobenius norm. This yields

$$\det(\mathcal{H}_{\text{approx}}) = D_{xx} D_{yy} - (0.9 D_{xy})^2.$$

Furthermore, the filter responses are normalised with respect to the r
This guarantees a constant Frobenius norm for any filter size.

Scale spaces are usually implemented as image pyramids. The in
repeatedly smoothed with a Gaussian and subsequently sub-sampled in
achieve a higher level of the pyramid. Due to the use of box filters and
images, we do not have to iteratively apply the same filter to the ou
previously filtered layer, but instead can apply such filters of any size a
the same speed directly on the original image, and even in parallel (alth
latter is not exploited here). Therefore, the scale space is analysed by u
the filter size rather than iteratively reducing the image size. The outp
above 9×9 filter is considered as the initial scale layer, to which we wi
scale $s = 1.2$ (corresponding to Gaussian derivatives with $\sigma = 1.2$). The
layers are obtained by filtering the image with gradually bigger mask
into account the discrete nature of integral images and the specific str
our filters. Specifically, this results in filters of size 9×9, 15×15, 21×21
etc. At larger scales, the step between consecutive filter sizes should
accordingly. Hence, for each new octave, the filter size increase is doubl
from 6 to 12 to 24). Simultaneously, the sampling intervals for the extr
the interest points can be doubled as well.

As the ratios of our filter layout remain constant after scaling, the
imated Gaussian derivatives scale accordingly. Thus, for example, our
filter corresponds to $\sigma = 3 \times 1.2 = 3.6 = s$. Furthermore, as the Froben
remains constant for our filters, they are already scale normalised [26]

In order to localise interest points in the image and over scales
maximum suppression in a $3 \times 3 \times 3$ neighbourhood is applied. The
of the determinant of the Hessian matrix are then interpolated in s

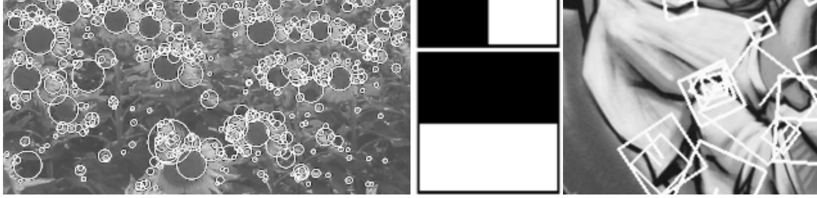

Fig. 2. Left: Detected interest points for a Sunflower field. This kind of sce
clearly the nature of the features from Hessian-based detectors. Middle: Ha
types used for SURF. Right: Detail of the Graffiti scene showing the size
scriptor window at different scales.

image space with the method proposed by Brown *et al.* [27]. Scale sp
polation is especially important in our case, as the difference in scale
the first layers of every octave is relatively large. Fig. 2 (left) shows an
of the detected interest points using our 'Fast-Hessian' detector.

4 SURF Descriptor

The good performance of SIFT compared to other descriptors [8] is re
Its mixing of crudely localised information and the distribution of gr
lated features seems to yield good distinctive power while fending off t
of localisation errors in terms of scale or space. Using relative stren
orientations of gradients reduces the effect of photometric changes.

The proposed SURF descriptor is based on similar properties, with a
ity stripped down even further. The first step consists of fixing a rep
orientation based on information from a circular region around the
point. Then, we construct a square region aligned to the selected ori
and extract the SURF descriptor from it. These two steps are now
in turn. Furthermore, we also propose an upright version of our descr
SURF) that is not invariant to image rotation and therefore faster
pute and better suited for applications where the camera remains mo
horizontal.

4.1 Orientation Assignment

In order to be invariant to rotation, we identify a reproducible orientati
interest points. For that purpose, we first calculate the Haar-wavelet
in x and y direction, shown in Fig. 2, and this in a circular neighbou
radius $6s$ around the interest point, with s the scale at which the inter
was detected. Also the sampling step is scale dependent and chosen to
keeping with the rest, also the wavelet responses are computed at tha

wavelets is $4s$.

Once the wavelet responses are calculated and weighted with a Gaus 2.5s) centered at the interest point, the responses are represented as ve space with the horizontal response strength along the abscissa and th response strength along the ordinate. The dominant orientation is esti calculating the sum of all responses within a sliding orientation window an angle of $\frac{\pi}{3}$. The horizontal and vertical responses within the wi summed. The two summed responses then yield a new vector. The lon vector lends its orientation to the interest point. The size of the sliding is a parameter, which has been chosen experimentally. Small sizes fire dominating wavelet responses, large sizes yield maxima in vector length not outspoken. Both result in an unstable orientation of the interest reg the U-SURF skips this step.

4.2 Descriptor Components

For the extraction of the descriptor, the first step consists of const square region centered around the interest point, and oriented along th tion selected in the previous section. For the upright version, this transf is not necessary. The size of this window is $20s$. Examples of such squar are illustrated in Fig. 2.

The region is split up regularly into smaller 4×4 square sub-regions. T important spatial information in. For each sub-region, we compute a fe features at 5×5 regularly spaced sample points. For reasons of simplicit d_x the Haar wavelet response in horizontal direction and d_y the Haa response in vertical direction (filter size $2s$). "Horizontal" and "verti is defined in relation to the selected interest point orientation. To inc robustness towards geometric deformations and localisation errors, the d_x and d_y are first weighted with a Gaussian ($\sigma = 3.3s$) centered at th point.

Then, the wavelet responses d_x and d_y are summed up over each s and form a first set of entries to the feature vector. In order to bri formation about the polarity of the intensity changes, we also extract of the absolute values of the responses, $|d_x|$ and $|d_y|$. Hence, each s has a four-dimensional descriptor vector $\mathbf{v}$ for its underlying intensity $\mathbf{v} = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|)$. This results in a descriptor vector fo sub-regions of length 64. The wavelet responses are invariant to a bias nation (offset). Invariance to contrast (a scale factor) is achieved by tu descriptor into a unit vector.

Fig. 3 shows the properties of the descriptor for three distinctively image intensity patterns within a subregion. One can imagine combin such local intensity patterns, resulting in a distinctive descriptor.

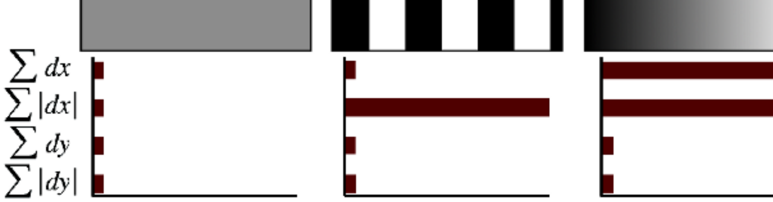

Fig. 3. The descriptor entries of a sub-region represent the nature of the u
intensity pattern. Left: In case of a homogeneous region, all values are relat
Middle: In presence of frequencies in x direction, the value of $\sum |d_x|$ is hig
others remain low. If the intensity is gradually increasing in x direction, bo
$\sum d_x$ and $\sum |d_x|$ are high.

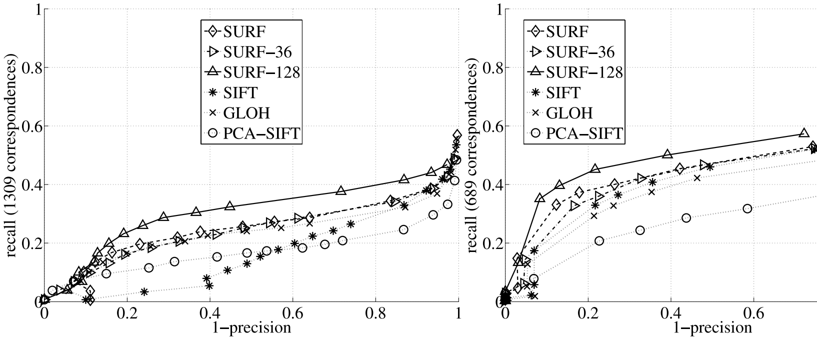

Fig. 4. The *recall* vs. *(1-precision)* graph for different binning methods and tw
matching strategies tested on the 'Graffiti' sequence (image 1 and 3) with a vi
of 30 degrees, compared to the current descriptors. The interest points are
with our 'Fast Hessian' detector. Note that the interest points are not affine
The results are therefore not comparable to the ones in [8]. SURF-128 co
to the extended descriptor. Left: Similarity-threshold-based matching strate
Nearest-neighbour-ratio matching strategy (See section 5).

In order to arrive at these SURF descriptors, we experimented w
and more wavelet features, using d_x^2 and d_y^2, higher-order wavelets, PCA
values, average values, etc. From a thorough evaluation, the proposed se
out to perform best. We then varied the number of sample points and sub
The 4×4 sub-region division solution provided the best results. Conside
subdivisions appeared to be less robust and would increase matching
much. On the other hand, the short descriptor with 3×3 subregions (S
performs worse, but allows for very fast matching and is still quite a
in comparison to other descriptors in the literature. Fig. 4 shows only
these comparison results (SURF-128 will be explained shortly).

separately for $d_y < 0$ and $d_y \geq 0$. Similarly, the sums of d_y and $|d_y|$ up according to the sign of d_x, thereby doubling the number of featu descriptor is more distinctive and not much slower to compute, but match due to its higher dimensionality.

In Figure 4, the parameter choices are compared for the standard scene, which is the most challenging of all the scenes in the evaluati Mikolajczyk [8], as it contains out-of-plane rotation, in-plane rotation brightness changes. The extended descriptor for 4×4 subregions (SU comes out to perform best. Also, SURF performs well and is faster t Both outperform the existing state-of-the-art.

For fast indexing during the matching stage, the sign of the Lapla the trace of the Hessian matrix) for the underlying interest point is Typically, the interest points are found at blob-type structures. Th the Laplacian distinguishes bright blobs on dark backgrounds from th situation. This feature is available at no extra computational cost, already computed during the detection phase. In the matching stage compare features if they have the same type of contrast. Hence, this information allows for faster matching and gives a slight increase in perf

5 Experimental Results

First, we present results on a standard evaluation set, fot both the det the descriptor. Next, we discuss results obtained in a real-life object re application. All detectors and descriptors in the comparison are base original implementations of authors.

Standard Evaluation. We tested our detector and descriptor using t sequences and testing software provided by Mikolajczyk [1]. These are real textured and structured scenes. Due to space limitations, we can the results on all sequences. For the detector comparison, we selected viewpoint changes (Graffiti and Wall), one zoom and rotation (Boat) and changes (Leuven) (see Fig. 6, discussed below). The descriptor evalua shown for all sequences except the Bark sequence (see Fig. 4 and 7).

For the detectors, we use the repeatability score, as described in indicates how many of the detected interest points are found in both relative to the lowest total number of interest points found (where only of the image that is visible in both images is taken into account).

The detector is compared to the difference of Gaussian (DoG) de Lowe [2], and the Harris- and Hessian-Laplace detectors proposed by jczyk [15]. The number of interest points found is on average very simi detectors. This holds for all images, including those from the databas

[1] http://www.robots.ox.ac.uk/~vgg/research/affine/

detector	threshold	nb of points	comp. time (msec)
Fast-Hessian	600	1418	120
Hessian-Laplace	1000	1979	650
Harris-Laplace	2500	1664	1800
DoG	default	1520	400

the object recognition experiment, see Table 1 for an example. As ca
our 'Fast-Hessian' detector is more than 3 times faster that DoG an
faster than Hessian-Laplace. At the same time, the repeatability for ou
is comparable (Graffiti, Leuven, Boats) or even better (Wall) than for
petitors. Note that the sequences Graffiti and Wall contain out-of-plane
resulting in affine deformations, while the detectors in the compariso
rotation- and scale invariant. Hence, these deformations have to be t
the overall robustness of the features.

The descriptors are evaluated using recall-(1-precision) graph
[4] and [8]. For each evaluation, we used the first and the fourth ima
sequence, except for the Graffiti (image 1 and 3) and the Wall scene
and 5), corresponding to a viewpoint change of 30 and 50 degrees, res
In figures 4 and 7, we compared our SURF descriptor to GLOH, SIFT a
SIFT, based on interest points detected with our 'Fast-Hessian' detect
outperformed the other descriptors for almost all the comparisons.
we compared the results using two different matching techniques, one
the similarity threshold and one based on the nearest neighbour rati
for a discussion on these techniques). This has an effect on the ranki
descriptors, yet SURF performed best in both cases. Due to space li
only results on similarity threshold based matching are shown in Fig.
technique is better suited to represent the distribution of the descri
feature space [8] and it is in more general use.

The SURF descriptor outperforms the other descriptors in a syste
significant way, with sometimes more than 10% improvement in reca
same level of precision. At the same time, it is fast to compute (see
The accurate version (SURF-128), presented in section 4, showed slig
ter results than the regular SURF, but is slower to match and ther
interesting for speed-dependent applications.

Table 2. Computation times for the joint detector - descriptor implementati
on the first image of the Graffiti sequence. The thresholds are adapted i
detect the same number of interest points for all methods. These relative
also representative for other images.

	U-SURF	SURF	SURF-128	SIFT
time (ms):	255	354	391	1036

Fig. 5. An example image from the reference set (left) and the test set (rig
the difference in viewpoint and colours.

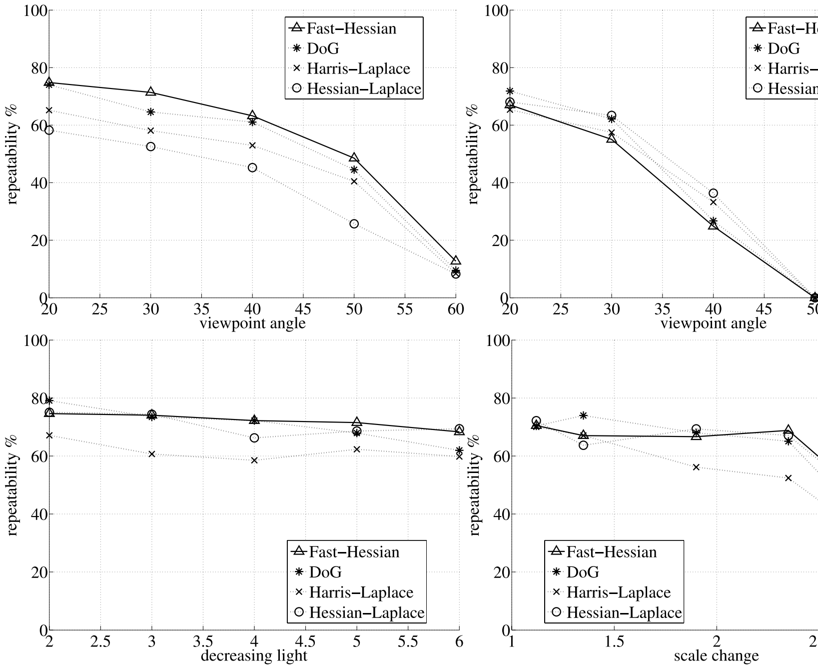

Fig. 6. Repeatability score for image sequences, from left to right and top t
Wall and Graffiti (Viewpoint Change), Leuven (Lighting Change) and Boat (
Rotation)

Note that throughout the paper, including the object recognition ex
we always use the same set of parameters and thresholds (see table
timings were evaluated on a standard Linux PC (Pentium IV, 3GHz).

Object Recognition. We also tested the new features on a practical ap
aimed at recognising objects of art in a museum. The database consis
images of 22 objects. The images of the test set (116 images) were t
der various conditions, including extreme lighting changes, objects in

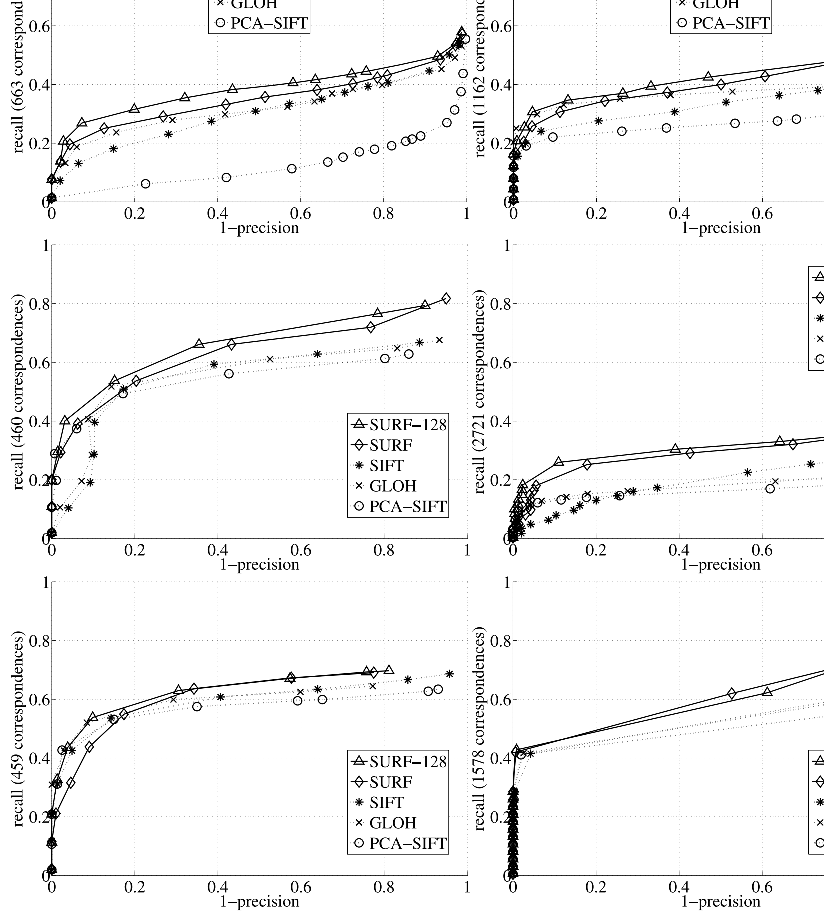

Fig. 7. Recall, 1-Precision graphs for, from left to right and top to botto
point change of 50 (Wall) degrees, scale factor 2 (Boat), image blur (Bikes a
brightness change (Leuven) and JPEG compression (Ubc)

glass cabinets, viewpoint changes, zoom, different camera qualities, e
over, the images are small (320×240) and therefore more challenging f
recognition, as many details get lost.

In order to recognise the objects from the database, we proceed as fol
images in the test set are compared to all images in the reference set by
their respective interest points. The object shown on the reference im
the highest number of matches with respect to the test image is chos
recognised object.

is closer than 0.7 times the distance of the second nearest neighbour. T
nearest neighbour ratio matching strategy [18, 2, 7]. Obviously, additiona
ric constraints reduce the impact of false positive matches, yet this can b
top of any matcher. For comparing reasons, this does not make sense, as t
hide shortcomings of the basic schemes. The average recognition rates r
results of our performance evaluation. The leader is SURF-128 with 85.7%
tion rate, followed by U-SURF (83.8%) and SURF (82.6%). The other de
achieve 78.3% (GLOH), 78.1% (SIFT) and 72.3% (PCA-SIFT).

6 Conclusion

We have presented a fast and performant interest point detection-de
scheme which outperforms the current state-of-the art, both in speed a
racy. The descriptor is easily extendable for the description of affine
regions. Future work will aim at optimising the code for additional spe
binary of the latest version is available on the internet[2].

Acknowledgements. The authors gratefully acknowledge the supp
Swiss SNF NCCR project IM2, Toyota-TME and the Flemish Fund for
Research.

References

1. Lindeberg, T.: Feature detection with automatic scale selection. IJC
(1998) 79 – 116
2. Lowe, D.: Distinctive image features from scale-invariant keypoints, casc
ing approach. IJCV **60** (2004) 91 – 110
3. Mikolajczyk, K., Schmid, C.: An affine invariant interest point detector. I
(2002) 128 – 142
4. Ke, Y., Sukthankar, R.: PCA-SIFT: A more distinctive representation
image descriptors. In: CVPR (2). (2004) 506 – 513
5. Tuytelaars, T., Van Gool, L.: Wide baseline stereo based on local, affinely
regions. In: BMVC. (2000) 412 – 422
6. Matas, J., Chum, O., M., U., Pajdla, T.: Robust wide baseline stereo fr
mally stable extremal regions. In: BMVC. (2002) 384 – 393
7. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descri
CVPR. Volume 2. (2003) 257 – 263
8. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descripto
27 (2005) 1615–1630
9. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J.
itzky, F., Kadir, T., Van Gool, L.: A comparison of affine region detect
65 (2005) 43–72

[2] http://www.vision.ee.ethz.ch/~surf/

ICCV. Volume I. (2001) 525 – 531
12. Lowe, D.: Object recognition from local scale-invariant features. In: ICC
13. Kadir, T., Brady, M.: Scale, saliency and image description. IJCV **45(**
83 – 105
14. Jurie, F., Schmid, C.: Scale-invariant shape features for recognition of c
egories. In: CVPR. Volume II. (2004) 90 – 96
15. Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point
IJCV **60** (2004) 63 – 86
16. Florack, L.M.J., Haar Romeny, B.M.t., Koenderink, J.J., Viergever, M.A
intensity transformations and differential invariants. JMIV **4** (1994) 171
17. Mindru, F., Tuytelaars, T., Van Gool, L., Moons, T.: Moment invariants
nition under changing viewpoint and illumination. CVIU **94** (2004) 3–2
18. Baumberg, A.: Reliable feature matching across widely separated views. I
(2000) 774 – 781
19. Schaffalitzky, F., Zisserman, A.: Multi-view matching for unordered ima
"How do I organize my holiday snaps?". In: ECCV. Volume 1. (2002) 4
20. Freeman, W.T., Adelson, E.H.: The design and use of steerable filters.
(1991) 891 – 906
21. Carneiro, G., Jepson, A.: Multi-scale phase-based local features. In: C
(2003) 736 – 743
22. Se, S., Ng, H., Jasiobedzki, P., Moyung, T.: Vision based modeling and
tion for planetary exploration rovers. Proceedings of International Ast
Congress (2004)
23. Viola, P., Jones, M.: Rapid object detection using a boosted cascade
features. In: CVPR (1). (2001) 511 – 518
24. Koenderink, J.: The structure of images. Biological Cybernetics **50** (19
370
25. Lindeberg, T.: Discrete Scale-Space Theory and the Scale-Space Prim
PhD, KTH Stockholm,. KTH (1991)
26. Lindeberg, T., Bretzner, L.: Real-time scale selection in hybrid multi-sc
sentations. In: Scale-Space. (2003) 148–163
27. Brown, M., Lowe, D.: Invariant features from interest point groups. In
(2002)