*Computer Vision Winter Workshop 2007, Michael Grabner, Helmut Grabner (eds.)*
*St. Lambrecht, Austria, February 6–8*
*Graz Technical University*

# Colour Interest Points for Image Retrieval

Julian Stöttinger[1], Nicu Sebe[2], Theo Gevers[2], and Allan Hanbury[1]

[1]PRIP, Institute of Computer Aided Automation,
Vienna University of Technology, Favoritenstraße 9/1832, A-1040 Vienna, Austria
stoettij@prip.tuwien.ac.at, hanbury@prip.tuwien.ac.at

[2]Faculty of Science, University of Amsterdam,
Kruislaan 403, 1098 SJ Amsterdam, The Netherlands
nicu@science.uva.nl, gevers@science.uva.nl

**Abstract** *In image retrieval scenarios, many methods use interest point detection at an early stage to find regions in which descriptors are calculated. Finding salient locations in image data is crucial for these tasks. Observing that most current methods use only the luminance information of the images, we investigate the use of colour information in interest point detection. Based on the Harris corner detector, a way to use multi-channel images is explored and different colour spaces are evaluated. To determine the characteristic scale of an interest point, a new colour scale selection method is presented. We show that using colour information and boosting salient colours results in improved performance in retrieval tasks.*

## 1 Introduction

Interest points in images are useful in a wide variety of applications, including stereo matching and object recognition. Corners have long been considered as useful interest points. Corner detection can be traced back to Moravec [15] who measured the average change of intensity by shifting a local window by a small amount in different directions. Harris and Stephens [3] improved the repeatability of Moravec detector under small image variations and near the edges. The Harris detector, in combination with a rotational invariant descriptor, was also used by Schmid and Mohr [16] when they extended local feature matching to general object recognition.

Lindeberg [6] proposed an "interesting scale level" detector which is based on determining maxima over scale of a normalized blob measure. The Laplacian-of-Gaussian (LoG) function is used for building the scale space. Mikolajczyk [11] showed that this function is very suitable for automatic scale selection of structures. An efficient algorithm for use in object recognition was proposed by Lowe [8]. This algorithm constructs a scale space pyramid using Difference-of-Gaussian (DoG) filters. The DoG are used to obtain an efficient approximation of the LoG. From the local 3D maxima a robust descriptor is built for matching purposes. The disadvantage of using DoG or LoG is that the repeatability of the extracted features is not optimal since both DoG and LoG not only respond to blobs, but also to high gradients in one direction. Because of this, the localization of the features may not be very accurate.

An approach that intuitively arises from this observation is the separation of the feature detector and the scale selection. The original Harris detector [3] is robust to noise and lighting variations, but only to a small extent to scale changes [17]. To deal with this Dufournoud et al. [1] proposed the scale adapted Harris operator. Given the scale adapted Harris operator, a scale space can be created. Local 3D maxima in this scale space can be taken as salient points. Mikolajczyk points out that the scale adapted Harris operator rarely attains a maximum over scales [11]. This results in very few points, which are not representative enough for capturing the image content. To address this problem, Mikolajczyk [11] proposed the Harris-Laplace detector that merges the scale-adapted Harris corner detector and the Laplacian based scale selection.

All the approaches presented above are intensity based. Since the luminance axis is the major axis of colour variation in the RGB colour cube, most interest points are found using just intensity. The additional colour based interest points might not dramatically increase the number of interest points. The distinctiveness of these colour based interest points is however much larger, and therefore colour can be of great importance when matching images. Furthermore, colour plays an important role in the pre-attentive stage in which features are detected. This means that the saliency value of a point also depends on the colour information that is present. Very relevant to our work is the research of van de Weijer and Gevers [19]. They aim at incorporating colour distinctiveness into the design of interest point detectors. In their work, the colour derivatives form the basis of a colour saliency boosting function since they are used in both the detection of the interest points, and the determination of the information content of the points. Furthermore, the histograms of colour image derivatives show distinctive statistical properties which are used in a colour saliency boosting function.

We propose a new method for the automatic determina-

tion of the characteristic scale of a region. Moreover, by using colour information in different colour spaces, the interest points gain distinctiveness and stability. The shifting of interest points towards colour differences leads to a different focus of interest points, which can be very useful in colorful images and natural, cluttered scenes.

We investigate different corner detection approaches and we evaluate them under varying circumstances including artificial well defined conditions and complex natural scenes.

## 2 Colour Corner Detection

In this section, we discuss the extension of the Harris corner detector to colour images, making use of colour spaces that are quasi-invariant to some variations in imaging conditions.

### 2.1 Colour Harris Corner Detector

The Harris corner detector introduced in [3] provides a cornerness measure for image data. It is calculated based on a second moment matrix $M$ describing the gradient distribution in the local neighbourhood of a point as

$$C_H(M) = \det(M) - \alpha \mathrm{trace}^2(M) \tag{1}$$

where the constant $\alpha$ indicates the slope of the "zero line".

A basic extension of the intensity based Harris detector is proposed by Montesinos et al. [14]. The second moment matrix they use is defined as

$$M = \mu(x,y,\sigma_I,\sigma_D) = \sigma_D^2 G(\sigma_I) \otimes \begin{bmatrix} R_x^2+G_x^2+B_x^2 & R_xR_y+G_xG_y+B_xB_y \\ R_xR_y+G_xG_y+B_xB_y & R_y^2+G_y^2+B_y^2 \end{bmatrix} \tag{2}$$

where $\otimes$ indicates convolution and the subscripts $x$ and $y$ indicate Gaussian derivatives at scale $\sigma_D$ in these directions. Instead of using just the intensity gradient, the gradient of each colour channel is determined. These values are summed and averaged using a Gaussian kernel $G$ with size $\sigma_I$. The choice of the colour channels is very important in extracting the Harris corners.

As already pointed out, the *RGB* colour space is highly correlated and therefore prone to illumination changes. However, in natural images, high contrast changes might take place. Therefore, a colour Harris detector in *RGB* colour space does not dramatically change the position of the corners compared to a luminance based approach. This is not the case for photos with low contrast changes as usually encountered in studio photography, artificial images, or highly postprocessed images. In Figures 1(a) and 2(a) we show the Harris energy and the extracted RGB corners. Note that the extracted points are spread all over image and do not concentrate on the object of interest.

Going from *RGB* to normalised *rgb*, the cornerness measurement (see Figure 1(b)) favours the colour changes between foreground and background and therefore, the silhouette of the parrot can be recognized in the plot. The main drawback of this colour space is that *rgb* is very unstable near zero illumination and the approach therefore leads to highly prominent corners extracted in the dark regions of the image. This is encountered in the lower right part of the image in Figure 1(b).

The second moment matrix can be computed using different colour models. The first step is to determine the gradients of each component of the *RGB* colour system. This is done using a convolution with the differentiation kernels of size $\sigma_D$. The gradients are then transformed into the desired colour system. By multiplication and summation of the transformed gradients, all components of the second moment matrix are computed. The values are averaged by a Gaussian integration kernel with size $\sigma_I$. Scale normalization is done again using a factor $\sigma_D^2$.

To write this procedure in symbolic form, we use a more general notation not restricted to one colour space. Colour space $C$ is used with its components $[c_1,\ldots,c_n]^T$, where $n$ is the number of colour system components. The second moment matrix is then

$$M = \mu(\mathbf{x},\sigma_I,\sigma_D) = \sigma_D^2 g(\sigma_I) \otimes \begin{bmatrix} L_x^2(\mathbf{x},\sigma_D) & L_xL_y(\mathbf{x},\sigma_D) \\ L_xL_y(\mathbf{x},\sigma_D) & L_y^2(\mathbf{x},\sigma_D) \end{bmatrix} \tag{3}$$

with the components $L_x^2$, $L_xL_y$ and $L_y^2$ defined as:

$$L_x^2(\mathbf{x},\sigma_D) = \sum_{i=1}^{n} c_{i,x}^2(\mathbf{x},\sigma_D) \tag{4}$$

$$L_xL_y(\mathbf{x},\sigma_D) = \sum_{i=1}^{n} c_{i,x}(\mathbf{x},\sigma_D) c_{i,y}(\mathbf{x},\sigma_D) \tag{5}$$

$$L_y^2(\mathbf{x},\sigma_D) = \sum_{i=1}^{n} c_{i,y}^2(\mathbf{x},\sigma_D) \tag{6}$$

where $c_{i,x}$ and $c_{i,y}$ denote the components of the transformed colour channel gradients, with $i \in [1,\ldots,n]$, and where the subscript $x$ or $y$ indicates the direction of the gradient.

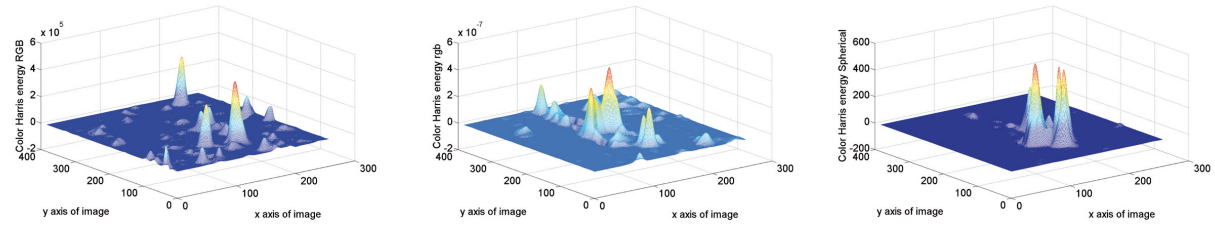
### 2.2 Quasi Invariant Colour Spaces

Due to the common photometric variations in imaging conditions such as shading, shadows, specularities and object reflectance, the components of the *RGB* colour system are correlated. By transforming the *RGB* colour coordinates to other systems, photometric alterations of features in images can be distinguished. For this purpose, we investigate different colour spaces: (1) the spherical colour space (Equation 7); (2) the opponent colour space *OCS* (Equation 8); and (3) the *HSI* colour space (Equation 9). In these decorrelated colour spaces only the photometric axes are influenced by the common photometric variations. In [19] the spatial derivatives are separated into photometric variant and invariant parts.

$$S = \begin{pmatrix} \theta \\ \varphi \\ r \end{pmatrix} = \begin{pmatrix} \tan^{-1}\left(\frac{G}{R}\right) \\ \sin^{-1}\left(\frac{\sqrt{R^2+G^2}}{\sqrt{R^2+G^2+B^2}}\right) \\ \sqrt{R^2+G^2+B^2} \end{pmatrix} \tag{7}$$

$$O = \begin{pmatrix} o_1 \\ o_2 \\ o_3 \end{pmatrix} = \begin{pmatrix} \frac{R-G}{\sqrt{2}} \\ \frac{R+G-2B}{\sqrt{6}} \\ \frac{R+G+B}{\sqrt{3}} \end{pmatrix} \tag{8}$$

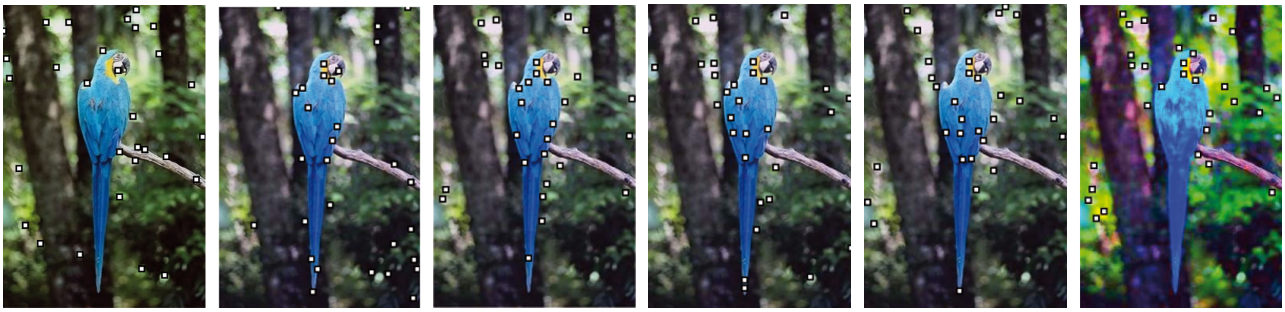
$$H = \begin{pmatrix} h \\ s \\ i \end{pmatrix} = \begin{pmatrix} \tan^{-1}\left(\frac{o_1}{o_2}\right) \\ \sqrt{o_1^2+o_2^2} \\ o_3 \end{pmatrix} \tag{9}$$




(a) *RGB* (b) *rgb* (c) Spherical colour space

(d) *OCS* (e) *HSI* (f) colour boosted *OCS*

**Figure 1:** Harris energy in different colour spaces. From left to right, the energy gets more and more distinctive and salient colours are favoured.



(a) *RGB* (b) *rgb* (c) Spherical colour space (d) *OCS* (e) *HSI* (f) colour boosted *OCS*

**Figure 2:** The 30 highest maxima extracted and visualized on the image for different colour spaces.

The spherical colour transformation (Equation (7)) has the shadow-shading direction as the $r$ coordinate. This overcomes the instability of *rgb* and is prone to decreasing the illumination of a colour. As shown in Figures 1(c) and 2(c), dark regions are more stable and there are fewer corners with shading changes involved. The orthonormal transformation into *OCS* (Equation (8)) provides specular variance. In Figures 1(d) and 2(d), the prioritization of the yellow - blue edge is shown. As this colour space is often described as simulating a primate's retinal processes, these opponent colours are on one axis and have therefore a large distance. The specular variance can also be seen at the left shoulder of the parrot in Figure 2(d).

A polar transformation on the first two axes of the *OCS* leads to the *HSI* colour space (Equation (9)). The derivative of the hue component is both shading and specular quasi-invariant [19]. The drawback is the ambiguity between black and white. Further, small changes around the gray axis result in large changes in the colour direction as can be seen in the upper left corner of Figure 2(e). As shown in Figure 1(e), colour changes between opponent colours are highly favoured. This leads to stable results under varying circumstances.

## 3 Scale Invariant Corner Detection

Using a fixed scale has one drawback: structures which are "too small" or "too large" are not taken into account. The goal is to develop a scale invariant description of corners in an image. In our retrieval context, this idea gives us the same locations, regardless of the size of the object in the image.

### 3.1 Scale Invariant Harris Corner Detection

The scale space of the Harris function is built by iteratively calculating the cornerness measurement $E$ under varying $\sigma_D$ and $\sigma_I$. As shown in several experiments [10, 11], the relation

$$\sigma_D = 3\sigma_I \tag{10}$$

performs best.

Using scale steps $s = 1,2,\ldots$ determining the iterations of the algorithm (typically between 8 and 20) with a factor $t$ of $\sigma_D$ from 1.2 to $\sqrt{2}$, the cornerness measurement $E$ is calculated as

$$E(x,y,s) = (x,y,s)M(x,y,t^s\sigma_D,\frac{t^s}{3}\sigma_I)\begin{pmatrix}x\\y\\s\end{pmatrix} \tag{11}$$

(a) Illumination (b) HSI

**Figure 3:** 30 extracted regions based on luminance and HSI information with $t = 1.2; s = 10; \sigma_I = 0.7$. The regions shift towards colour differences, specular, and shading changes are not regarded anymore. The parrot is therefore highly prioritized.

The second moment matrix $M$ is not changed at all. The amount of scale change is chosen by the need for preciseness of the corner location. In reasonably large objects in a retrieval context, $t = \sqrt{2}$ showed to be precise enough and is therefore used in these experiments. For more precise results, a value of $t \approx 1.2$ is recommended.

The number of steps is a crucial matter for the processing time. Each step has to be calculated on its own (but independently in parallel) and the processing time increases with the size of the kernels. Building this scale space of the Harris energy leads to a pyramid of the cornerness measurement $E$. In several other implementations of a scale invariant corner detector (e.g. [9, 18]) the Harris energy of scale levels is scale normalized by the factor $\sigma_I^2$ because the values tend to get lower as the scale gets higher.

The next step is to choose the characteristic structure. In this research, the Laplacian of Gaussian function $\Lambda$ is used to select the characteristic structure automatically [7, 9]. Extending it to the scale space chosen previously, the scale decision measurement $\Lambda$ is defined as

$$\Lambda(x,y,\sigma_D) = -\frac{1}{\pi(t^s\sigma_D)^4}\left(1-\frac{x^2+y^2}{1(t^s\sigma_D)^2}\right)e^{-\frac{x^2+y^2}{2(t^s\sigma_D)^2}} \otimes c_{u,v} \tag{12}$$

To make the maxima more stable, a raised cosine kernel is used to smooth the resulting data

$$c_{u,v} = \frac{1+((\frac{1}{2}-cos(\pi u))+(\frac{1}{2}-cos(\pi v)))}{3} \tag{13}$$

As suggested in [5], this kernel gives smoother borders than the Gaussian Kernel for scale decision.

A characteristic scale of a possible region is found if both the Harris Energy and the Laplacian of Gaussian are an extremum

$$\nabla\Lambda(x,y,s(\sigma_D)) = \nabla M(x,y,\sigma_I,\sigma_D) = \overline{0} \tag{14}$$

where $\overline{0}$ is the null vector. Results are shown in Figure 3. With this non-maxima suppression, the majority of data is discarded leaving $\widehat{E}$ and $\widehat{\Lambda}$ for which Equation (14) holds. Aiming for just one region per location and a reasonable distribution of regions over the input image, the following decision criterion was shown to perform best

$$\widehat{R}(x,y) = \begin{pmatrix} \max\left(\widehat{E}(x,y,*)\right) \\ 3t^{\arg\max\left(\widehat{\Lambda}(x,y,*)\right)}\sigma_D \end{pmatrix} \tag{15}$$

This leads to the function $\widehat{R}(x,y)$ defining all interest point candidates and the corresponding region size.

Within this step, there is not one location chosen, but a region of interest with a centered interest point as visualized in Figure 3. The information of this area is then used in the calculation of the descriptor (e.g. the SIFT descriptor).

### 3.2 Colored Scale Invariant Harris Corner Detection

In this section, we propose a method for including colour information in the scale decision. The input image is transformed to the same colour space as is used for the extraction of the Harris energy. After that, the image is globally analyzed: a principal component analysis (PCA) takes place to reduce the 3 colour dimensions of the input image to a one dimensional dataset $\widehat{I}(x,y)$, where

$$\widehat{I}(x,y) = \sqrt{3}\nu_\lambda I(x,y)^T \otimes c_{u,v} \tag{16}$$

by calculating the dot product of the colour information $I(x,y)$ and the corresponding eigenvector $\nu_\lambda$. For implementation reasons, the result is scaled back by $\sqrt{3}$. The training set is the whole image.

This analysis leads to a transformed one dimensional function which includes many of the advantages of the corresponding colour space, as described in [18]. Based on $\widehat{I}(x,y)$, Equation (12) can be applied and the characteristic scale can be chosen using the procedure described in Section 3.1.

Considering that the discrimination vector is chosen as the maximum of the sum of the distances between the values, the PCA, as the basis for the scale decision criterion, ensures that a trade-off between favouring rare colours and retaining information on similar colours is realized. Therefore, it can be seen as a relaxed colour boosting function within the dimension reduction. If salient values have larger distances than many others, less salient colours are disregarded and get similar values. If the distance to the rarest colours is not large enough, the transformation favours common colours. This transformation tends to lose less distance information than other transformations $f : \mathbb{R}^3 \to \mathbb{R}^1$ e.g. the one usually used by the luminance transform.

### 3.3 Colour Statistics and Boosting

A form of a general *saliency implies rarity* [4] approach for weighting colours is used for boosting interest points.

As proposed in [20], colours have different occurrence probabilities $p(\mathbf{v})$ and therefore different information content $I(\mathbf{v})$

$$I(\mathbf{v}) = -log(p(\mathbf{v})) \tag{17}$$

The idea is to boost rare colours to have higher saliency in the cornerness measurement. When looking for rare

colours, statistics made on the Corel Database containing 40 000 colour images showed that the three dimensional colour distribution was remarkably significant. For all considered colour spaces, one coordinate coincides with the axis of maximum variation (i.e., the luminance).

Traditionally, the derivatives of colour vectors with equal vector norms have equal impact on the saliency function. We wish to find a boosting function so that colour vectors having equal information content have equal impact on the saliency function. This is a *colour saliency boosting* transformation $g : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ such that

$$p(\mathbf{f}_x) = p(\mathbf{f}'_x) \leftrightarrow \|g(\mathbf{f}_x)\| = \|g(\mathbf{f}'_x)\| \tag{18}$$

where $\mathbf{f}_x$ and $\mathbf{f}'_x$ are the derivatives of two arbitrary colour coordinate vectors $\mathbf{f}$ and $\mathbf{f}'$. The transformation is obtained by deriving a function describing the surface of the 3 dimensional colour distribution, which can be approximated by an ellipsoid. The third coordinate of the colour space is already aligned with the luminance, which forms the longest axis of the ellipsoid. The other two axes are rotated so that they are aligned with the other two axes of the ellipsoid.

These derivative histograms can then be approximated by ellipsoids having the definition

$$(\alpha h_x^1)^2 + (\beta h_x^2)^2 + (\gamma h_x^3)^2 = R^2 \tag{19}$$

where $h_x^{[1..3]}$ is the transformation of a colour derivative to one of the colour spaces given in Equations (7)–(9) followed by the rotation to align the axes with those of the ellipsoid in the corresponding colour space. To find the transformation in Equation (18), the ellipsoid is transformed to a sphere, so that vectors of equal saliency lead to vectors of equal length. The function $g$ is therefore defined as

$$g(\mathbf{f}_x) = \Lambda h(\mathbf{f}_x) \tag{20}$$

which leads to a saliency boosting factor for each component of the corresponding colour space. For the opponent colour space, the diagonal matrix $\Lambda$ is given by

$$\Lambda = \begin{bmatrix} 0.850 & 0 & 0 \\ 0 & 0.524 & 0 \\ 0 & 0 & 0.065 \end{bmatrix} \tag{21}$$

The idea is now that these factors not only hold for the analyzed images, but for other natural images as well. Figures 1(f) and 2(f) show the Harris energy and the corresponding corners obtained by using the colour boosting transformation.

# 4 Results

In Section 4.1, visualizations of the performance of the different colour spaces are given. In Section 4.2, the repeatability under viewpoint transformations is compared to the existing state of the art implementations of [11] and the results are discussed. Finally, a retrieval scenario is studied in Section 4.3.

The latter two experiments use the Amsterdam Library of Object Images (ALOI)[1], which provides images of 1000 objects under supervised, predefined conditions on a dark background. Several transformations are precisely applied, including viewpoint transformation and varying light conditions [2].


(a) front view (b) 25° turned (c) 50° turned.

**Figure 5:** ALOI object number 46 under viewpoint transformations.

## 4.1 Shifting Interest Points Towards Colour

Taking colour information into account leads to a different definition of interest points. Every colour transformation provides other properties. In the case of the quasi invariant colour space, colour only is regarded. The colour only interest points consider only changes in colour, not in illumination, specular, or shadow changes. It is more likely to describe meaningful objects in real circumstances that way, as the results are not changed by different lighting conditions. Working with natural cluttered animal images, this helps to overcome one of the major problem: every natural image is taken in completely different circumstances, and therefore under completely different lighting conditions. Figures 4(a)–4(f) show the shifting of interest points towards the colour edges as more colour specific interest point detectors are used. Note that the background of Figure 4(f) is not disregarded, but the colour edges have higher priority. As the corner measurement expands heavily under the quasi invariant colour space, the Harris energy has a wider range of data. Taking more maxima into account, a stable and meaningful distribution of the interest points can be achieved.

## 4.2 Repeatability Experiment

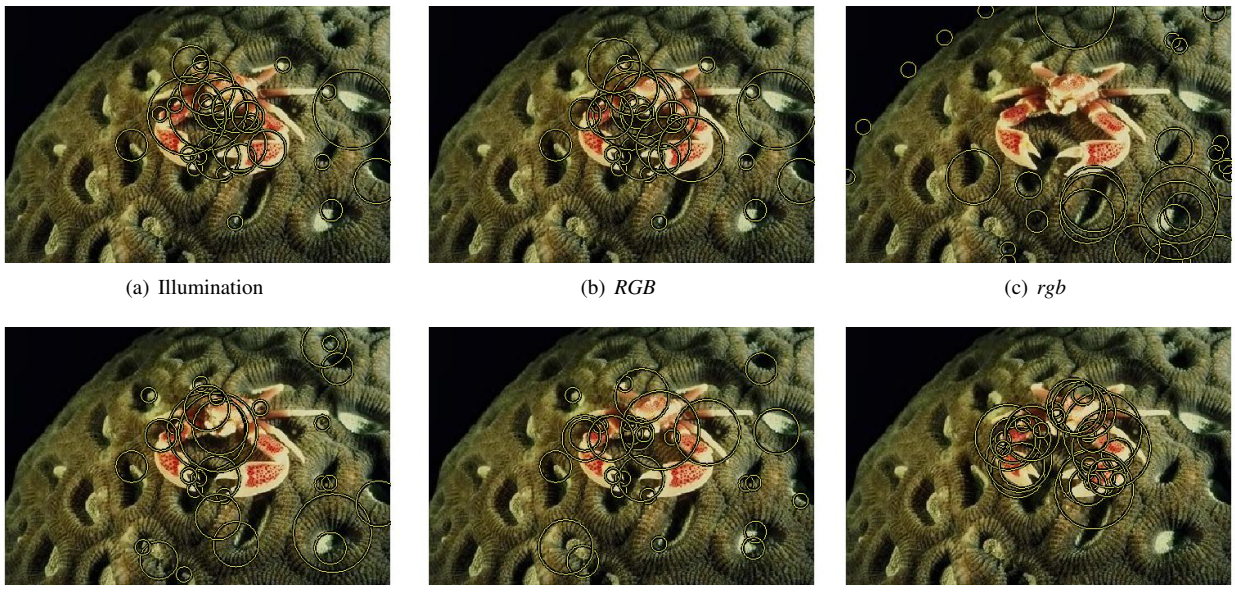
The experimental setup is the following: a colorful, flat object such as the one shown in Figure 5(a) is turned on a rotation stage in steps of 5° . Rotations of 50° in both directions are used. The images are captured at a distance of 124.5 cm to the center of the rotational stage. Light is provided by Halogen lights and a 48 mm camera lens is used. The input images are full colour 24 bit at a resolution of $768 \times 576$ and are in png format.

Mikolajczyk provides his interest point detectors in freely available binaries[2]. He also provides software to estimate the homography matrix and to evaluate repeatability performance. We used these binaries in order to compare our colour based system with the intensity based approaches.

The performance is measured by the repeatability rate, which is the percentage of corresponding points detected in two images. The higher the repeatability rate between two images, the more points can potentially be matched and the better are the matching and recognition results. A match is counted if the transformation of one image to the other

[1] http://staff.science.uva.nl/~aloi/

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



(a) Illumination (b) *RGB* (c) *rgb*

(d) *OCS* (e) colour boosted *OCS* (f) Quasi invariant *HSI*

**Figure 4:** Scale invariant interesting point extraction for different colour spaces. The background is highly structured with high illumination changes. The parameters are $t = 1.2; s = 10; \sigma_I = 0.7$.

one using the provided homography matrix leads to interest regions overlapping by more than 40%.

The experiments with these binaries are previously done with a test set of images of natural scenes [13]. Compared to the precise transformations of the ALOI database, the transformations are done in larger and looser steps. Therefore, the results of these experiments give higher repeatability rates than those with other datasets (e.g [12, 10]). Having simple, precise transformations on flat objects provides a more stable repeatability rate than previous experiments.

The following algorithms are tested:

- the *Quasi Invariant ScIv Harris* uses the approach under the quasi invariant *HSI* colour space.
- The *Harris Laplacian* corner detector is the scale invariant approach from the Mikolajczyk implementation. It is a Harris implementation using LoG for the scale determination.
- *Harris Affine* is the extension of this approach, using the results of the Harris Laplacian algorithm to detect the affine transformation of the region (Mikolajczyk's implementation is used for this algorithm too).
- *Colour boosted ScIv Harris* uses our approach under the opponent colour space and the colour statistics to boost rare colours.
- *RGB ScIv Harris* uses *RGB* information only.

As shown in Figure 6, the Harris Laplacian detector performs steadily about 5% better than the Harris Affine detector, a result which is explainable by the repeatability criteria. Both approaches use the same locations, as the majority of the algorithm is the same. Just the final stage of converting the scale invariant regions into affine invariant regions diminishes the area of the region by the direction of the gradients. These smaller, elliptic regions have smaller overlapping areas after transforming them back according to the given homography matrix. However, the result remains relatively stable beginning at over 70% and ending below 70% after the 50° transformation.

Performing this experiment in the *RGB* colour space, the results are quite similar to the luminance only approaches, until the transformation reaches a level of 35°. From this point, all colour based approaches perform better than those using only luminance information. Apparently, colour edges remain more stable under these transformations.

Using colour statistics (Section 3.3) and the *OCS* colour space, the salient colour differences become more distinct, and therefore the results become better. A drawback of this method is the instability to aliasing effects of the transformation, as seen in the 35° transformation.

The quasi invariant colour space performs best, as this approach takes only colour differences into account. Many parts of the image, like the blue shading, are disregarded completely, and only the colour changes of stable objects are taken as interest points. This leads to a 95% repeatability rate at a 30° transformation and an 85% rate after the full 50° transformation.

### 4.3 Image Retrieval

For the retrieval experiment, the impact of the extraction of interest points in a retrieval scenario is examined.

The retrieval scenario consists of 1000 objects captured as described in Section 4.2. For every object, 9 images are taken rotating the object 60° in both directions. From 5° to 30° and 355° to 330° rotation, the steps are taken in 5° increments. Up to 60° and 300°, respectively, the steps are carried out in 10° increments. This results in a database of 18000 images. Since the ALOI database delivers images of ob-

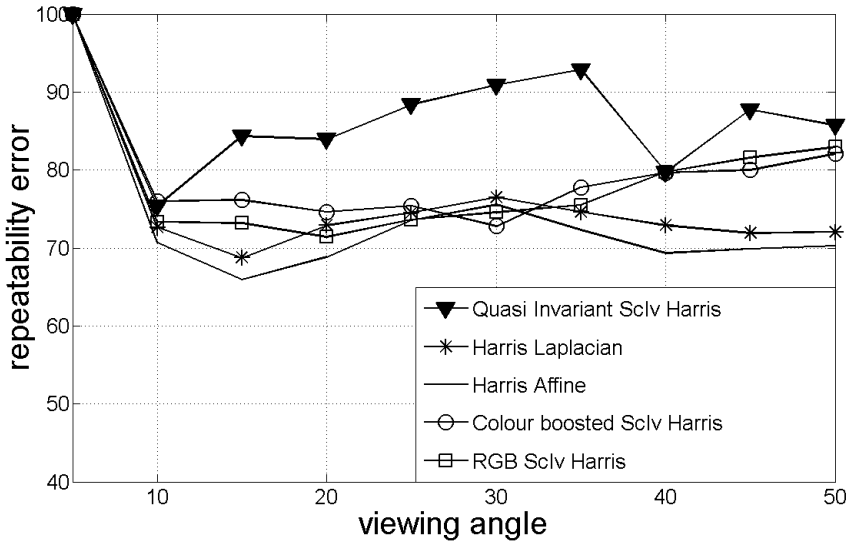



**Figure 6:** Repeatability experiment with ALOI database on full size resolution. Colour information gains stability under viewpoint transformation on colorful, flat objects.

jects on a dark background and image masks to completely disregard the background, the retrieval is obtained by object characteristics only. Query images are captured from the front view, the position which was omitted in the database. Every query image is processed as described above, except that no mask is applied to it. Therefore, it is possible that descriptors located in the background can occur in the query image.

We tested four algorithms for finding interest points:

- the *Illuminance Harris* indicates the scale invariant Harris approach from the reference implementation by Mikolajczyk.
- *DoG* uses the original SIFT binaries from Lowe[3] for the interest point extraction.
- The *OCS Harris* defines our scale invariant approach in the opponent colour space
- the *cb OCS Harris* boosts the opponent colour space with the colour statistics and extracts scale invariant interest points based on the weighted colours.

All these scale invariant interest points provide the locations for the calculation of the SIFT descriptors (also obtained using the implementation by Mikolajczyk). Therefore, the only difference between the four different retrieval tests is in the interest point extraction stage.

The difference between two images is determined by first calculating the Euclidean distances between each possible pair of (normalised) descriptors. The mean of the $N$ smallest distances is then taken to be the distance between the images (we use $N = 100$). As the retrieval performance measurement, the precision and recall values are calculated for the 30 best matches to the query image.

[3] `http://www.cs.ubc.ca/~lowe/keypoints/`

As shown in Figure 7, the overall performance of the approach based on the colour boosted OCS is better than the two illumination based methods. The DoG in the original SIFT implementation outperforms the illuminance Harris, especially in low contrast, dark images. The applied colour boosting factors improve the ability to describe an object under heavy viewpoint changes, as the performance is steadily better than in standard OCS.

The black background provides predominantly dark images with constant lighting. From this point of view, the illumination only based methods should not have problems in stability in changing contrast. In another context, the illumination based methods suffer in great instability: when the rotating object moves a surface away from the light source, the illumination on this surface gets less and less as there is almost no ambient light. In these cases, the colour boosted OCS based method performs better, as the colours get more distinctive, especially under dark circumstances. Illumination based approaches tend to lose meaningful locations under these illumination changes.

## 5 Conclusion

Using colour distances for corner measurement can shift the interest points to more meaningful, stable and distinct locations than luminance based methods. A colour scale selection leads to a better stability under transformations. Both the corner measurement and the scale selection can be transformed into various colour spaces, and we can take advantage of different properties of these transformations. Using correlated colour, boosted colour or colour invariant information, the method gains performance over luminance based methods. In retrieval scenarios, our approach was shown to be more distinct and stable, which leads to a higher and more precise retrieval rate than reference implementations.

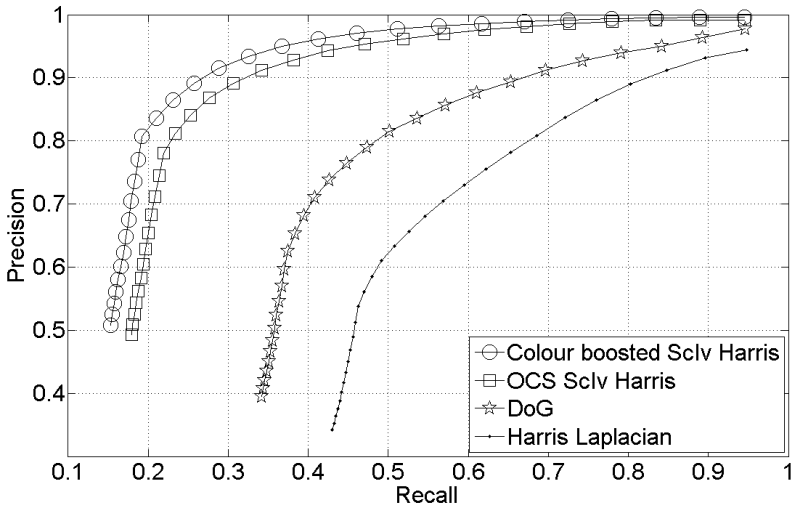



**Figure 7:** Precision recall graph of the retrieval experiment on the whole ALOI database. All SIFT descriptors are calculated by the Mikolajczyk binary, differences in performance are caused by different interesting regions only.

## Acknowledgement

This work was partly supported by the Austrian Science Foundation (FWF) under grant SESAME (P17189-N04), and the European Union Network of Excellence MUSCLE (FP6-507752).

## References

[1] Y. Dufournaud, C. Schmid, and R. Horaud. Matching images with different resolutions. In *CVPR*, pages 612–618, 2000.

[2] J. M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders. The Amsterdam library of object images. *Int. J. Comput. Vision*, 61(1):103–112, 2005.

[3] Chris Harris and Mike Stephens. A combined corner and edge detector. In *Proceedings 4th Alvey Visual Conference, UK*, 1988.

[4] Timor Kadir and Michael Brady. Saliency, scale and image description. *Int. J. Comput. Vision*, 45(2):83–105, 2001.

[5] C. S. Kenney, M. Zuliani, and B. S. Manjunath. An axiomatic approach to corner detection. In *Proc. CVPR*, volume 1, pages 191–197, 2005.

[6] T. Lindeberg. Feature detection with automatic scale selection. *IJCV*, 30(2):79–116, 1998.

[7] T. Lindeberg and J. Garding. Shape-adapted smoothing in estimation of 3-d shape cues from affine deformations of local 2-d brightness structure, 1997.

[8] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, November 2004.

[9] Krystian Mikolajczyk and Cordelia Schmid. Indexing based on scale invariant interest points. In *Proceedings of the 8th International Conference on Computer Vision*, pages 525–531, 2001.

[10] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. In *ECCV (1)*, pages 128–142, 2002.

[11] Krystian Mikolajczyk and Cordelia Schmid. Scale and affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.

[12] Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(10):1615–1630, 2005.

[13] Krystian Mikolajczyk, Tinne Tuytelaars, Cordelia Schmid, Andrew Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1/2):43–72, 2005.

[14] P. Montesinos, V. Gouet, and R. Deriche. Differential invariants for color images, 1998.

[15] Hans Moravec. Obstacle avoidance and navigation in the real world by a seeing robot rover. In *tech. report CMU-RI-TR-80-03, Robotics Institute, Carnegie Mellon University, doctoral dissertation, Stanford University*. September 1980.

[16] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(5):530–535, 1997.

[17] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *IJCV*, 37(2):151–172, 2000.

[18] Nicu Sebe, Theo Gevers, Sietse Dijkstra, and Joost van de Weijer. Evaluation of intensity and color corner detectors for affine invariant salient regions. In *CVPRW '06: Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop*, page 18, Washington, DC, USA, 2006. IEEE Computer Society.

[19] Joost van de Weijer and Theo Gevers. Edge and corner detection by photometric quasi-invariants. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 27(4):625–630, 2005.

[20] Joost van de Weijer, Theo Gevers, and Andrew Bagdanov. Boosting color saliency in image feature detection. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 28(1):150–156, jan 2006.