DO COLOUR INTEREST POINTS IMPROVE IMAGE RETRIEVAL?

J. Stoettinger, A. Hanbury

PRIP, Institute of Computer Aided Automation,
Vienna University of Technology, Austria

N. Sebe, T. Gevers

Intelligent Sensory Information Systems,
University of Amsterdam, The Netherlands

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. A way to use multi-channel information in the Harris corner detector 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.

Index Terms— Colour image analysis, Colour interest points, Image retrieval, Object recognition

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. The classical corner detector [1] was used in combination with a rotational invariant descriptor in [2] to extend local feature matching to general object recognition. Lindeberg [3] used the Laplacian-of-Gaussian (LoG) function for building the scale space in his "interesting scale level" detector. Mikolajczyk [4] showed that this function is very suitable for automatic scale selection of structures.

The original Harris detector [1] is robust to noise and lighting variations, but only to a small extent to scale changes. To deal with this, Dufournoud et al. [5] proposed the scale adapted Harris operator. Mikolajczyk [4] 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. The distinctiveness of colour based interest points is however much larger, and therefore colour can be of great importance when matching images. Very relevant to our work is that of van de Weijer [6], who aims at incorporating colour distinctiveness into the design of interest point detectors.

The main contribution of this paper is a new colour-based method for the automatic selection of the characteristic scale of an image region. We also investigate the use of various colour spaces for interest point detection and scale determination. The shifting of interest points towards colour differences can be useful in colorful images and natural, cluttered scenes. We provide a detailed evaluation of the use of these techniques in image retrieval.

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

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 [1] provides a cornerness measure for image data, calculated based on a second moment matrix M describing the gradient distribution in the local neighbourhood of a point as $C_H(M) = \det(M) - \alpha \text{trace}^2(M)$ where α indicates the slope of the "zero line".

An extension of the Harris detector to colour is proposed in [7]. Their second moment matrix is defined as

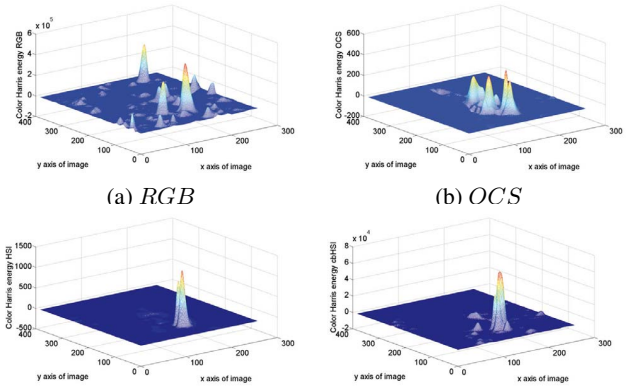
$$M = \sigma_D^2 G(\sigma_I) \otimes \begin{bmatrix} R_x^2 + G_x^2 + B_x^2 & R_x R_y + G_x G_y + B_x B_y \\ R_x R_y + G_x G_y + B_x B_y & R_y^2 + G_y^2 + B_y^2 \end{bmatrix} \quad (1)$$

where $\otimes$ indicates convolution and the subscripts x and y indicate Gaussian derivatives at scale σ_D in these directions. σ_I is the integration scale. The second moment matrix can be computed using different colour models. The first step is to determine the gradients of each color component and then the gradients are transformed into the desired colour system.

Because of common photometric variations in imaging conditions such as shading, shadows, specularities and object reflectance, the components of the RGB colour system are correlated and therefore sensitive to illumination changes. However, in natural images, high contrast changes may appear. Therefore, a colour Harris detector in RGB space does not dramatically change the position of the corners compared to a luminance based approach (see Figs. 1(a) and 2(a)). Nomalized rgb overcomes the correlation of RGB and favours colour changes. The main drawback, however, is its instability in dark regions (see lower right region of Fig. 2(b)). We can overcome this by using quasi invariant colour spaces.

2.2. Quasi Invariant Colour Spaces

By transforming the RGB colour coordinates to other systems, photometric alterations of features in images can be distinguished. For this purpose, we investigate: (1) the opponent colour space OCS (Eq. 2); and (2) the HSI colour space (Eq. 3) [6].



(a) RGB (b) OCS

(c) HSI (d) colour boosted OCS

Fig. 1. Harris energy in different colour spaces for the parrot image in Fig. 2. From (a) to (d), salient colors become more favoured.

$$O=\begin{pmatrix}o_1\\o_2\\o_3\end{pmatrix}=\begin{pmatrix}\frac{R-G}{\sqrt{2}}\\\frac{R+G-2B}{\sqrt{6}}\\\frac{R+G+B}{\sqrt{3}}\end{pmatrix} \quad (2) \qquad H=\begin{pmatrix}h\\s\\i\end{pmatrix}=\begin{pmatrix}\tan^{-1}\left(\frac{o_1}{o_2}\right)\\\sqrt{o_1^2+o_2^2}\\o_3\end{pmatrix} \quad (3)$$

Focusing more on colour differences, Figs. 1(b) and 2(c) show a higher Harris energy inside the parrot, even at dark blue/brown edges. The HSI colour space overcomes shading, shadows, specularities and object reflectance, therefore only colour differences are taken into account (see Figs. 1(c) and 2(d)).

2.3. Colour Statistics and Boosting

As proposed in [9], colours have different occurrence probabilities $p(\mathbf{v})$ and therefore different information content $I(\mathbf{v})$:

$$I(\mathbf{v}) = -log(p(\mathbf{v})) \quad (4)$$

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

$$p(\mathbf{f}_x) = p(\mathbf{f}'_x) \leftrightarrow \|g(\mathbf{f}_x)\| = \|g(\mathbf{f}'_x)\| \quad (5)$$

where $\mathbf{f}_x$ and $\mathbf{f}'_x$ are the spatial derivatives of two arbitrary colour coordinate vectors $\mathbf{f}$ and $\mathbf{f}'$.

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

$$(\alpha h_x^1)^2 + (\beta h_x^2)^2 + (\gamma h_x^3)^2 = R^2 \quad (6)$$

where $h_x^{[1..3]}$ is the transformation of a colour derivative to one of the colour spaces given in Eqs. (2)–(3) followed by the rotation to align the axes with those of the ellipsoid in the corresponding colour space. To find the transformation in Eq. (5), 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

$$g(\mathbf{f}_x) = \Lambda h(\mathbf{f}_x) \quad (7)$$

which leads to a saliency boosting factor for each component of the corresponding colour space. For the opponent colour space, the diagonal matrix Λ is chosen as in [9]. Figs. 1(d) and 2(e) show the Harris energy and the corresponding corners obtained by using the color boosting transformation.

3. SCALE INVARIANT CORNER DETECTION

Using a fixed scale has the drawback that structures which are "too small" or "too large" are not taken into account. Hence, our aim is to develop a scale invariant description of corners in an image. In our retrieval context, this gives the same interest point 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, where

$$E(x, y, s) = (x, y)M\left(x, y, t^s\sigma_D, \frac{t^s}{3}\sigma_I\right)\begin{pmatrix}x\\y\end{pmatrix} \quad (8)$$

under varying σ_D and σ_I. As shown in several experiments [4], the relation $\sigma_D = 3\sigma_I$ performs best. We use scale steps $s = 1, 2, \ldots$ determining the iterations of the algorithm (typically between 8 and 20) with a factor t from 1.2 to $\sqrt{2}$. The amount of scale change is chosen by the need for preciseness of the corner location.

The next step is to choose the characteristic structure. In this research, the LoG function Λ (see [4]) is used to select the characteristic structure automatically. Extending it to the scale space chosen previously, the scale decision measurement Λ is

$$\Lambda(x, y, \sigma_D) = -\frac{1}{\pi(t^s\sigma_D)^4}\left(1 - \frac{x^2+y^2}{1(t^s\sigma_D)^2}\right)e^{-\frac{x^2+y^2}{2(t^s\sigma_D)^2}} \otimes c_{u,v} \quad (9)$$

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

$$c_{u,v} = \frac{1 + ((\frac{1}{2} - cos(\pi u)) + (\frac{1}{2} - cos(\pi v)))}{3} \quad (10)$$

As suggested in [8], 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 LoG are extrema

$$\nabla\Lambda(x, y, s(\sigma_D)) = \nabla M(x, y, \sigma_I, \sigma_D) = \overline{0} \quad (11)$$

where $\overline{0}$ is the null vector.

With this non-maxima suppression, the majority of data is discarded leaving $\widehat{E}$ and $\widehat{\Lambda}$ for which Eq. (11) 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

$$\widehat{R}(x, y) = \begin{pmatrix}\max\left(\widehat{E}(x, y, *)\right)\\3t^{\arg\max\left(\widehat{\Lambda}(x,y,*)\right)}\sigma_D\end{pmatrix} \quad (12)$$

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

(a) RGB (b) rgb (c) OCS (d) HSI (e) cb OCS (f) Illumination (g) HSI

Fig. 2. (a) – (e): The 30 highest maxima of the Harris energy extracted for different color spaces. (f) – (g): 30 extracted regions based on (f) luminance and (g) HSI, with $t = 1, 2; s = 10; \sigma_I = 0.7$. The regions shift towards colour differences. Specular and shading changes are not regarded. The parrot is therefore highly prioritized.

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, a principal component analysis (PCA) takes place to reduce the three colour dimensions of the input image to a one dimensional dataset $\widehat{I}(x, y)$ by calculating the dot product of the colour information $I(x, y)$ and the eigenvector ν_λ with the largest eigenvalue:

$$\widehat{I}(x, y) = \sqrt{3}\nu_\lambda I(x, y)^T \otimes c_{u,v} \tag{13}$$

This analysis leads to a transformed one dimensional function which includes many of the advantages of the corresponding colour space. Based on $\widehat{I}(x, y)$, Eq. (9) can be applied and the characteristic scale can be chosen using the procedure described in Sec. 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. Figs. 2(f) and (g) give a comparison of extracted regions using intensity only and colour information. Hence 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 $f : \mathbb{R}^3 \rightarrow \mathbb{R}^1$ e.g. the one usually used by the luminance transform.

4. RESULTS

We performed experiments on the ALOI database [10], which provides images of 1000 objects under supervised, predefined conditions on a dark background. The following algorithms are tested: The *Harris Laplacian* corner detector is the scale invariant approach from the Mikolajczyk implementation[1]. *DoG* stands for the extraction of interest points with the difference of Gaussians using Lowe's implementation[2]. *Harris Affine* is the extension of this approach, using the results of the Harris Laplacian algorithm to detect the affine transformation of the region (also using Mikolajczyk's implementation). The *Quasi Invariant ScIv Harris* uses our approach in the Quasi Invariant HSI colour space, and *Colour boosted ScIv Harris* uses the opponent colour space and colour statistics to boost rare colours. *RGB ScIv Harris* uses RGB information only.

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

[2] www.cs.ubc.ca/~lowe/keypoints/

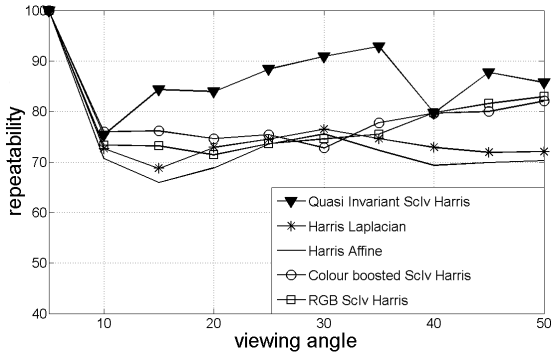


Fig. 3. Repeatability experiment with ALOI database.

4.1. Repeatability Experiment

In this experiment, we consider the objects rotated in steps of $5°$ with a range of $50°$ in both directions. The performance is measured by the repeatability, which is the percentage of corresponding regions detected in two images. A match is counted if the transformation of one image to the other with the provided homography matrix leads to interest regions overlapping by more than 40% [11].

As shown in Fig. 3, the Harris Laplacian detector performs steadily about 5% better than the Harris Affine detector, a result which is explainable by the repeatability criteria. However, the result remains relatively stable around 70%. When 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 (Sec. 2.3) and the OCS colour space, the salient colour differences become more distinct, and therefore the results improve. A drawback of this method is the instability due to aliasing effects of the transformation,

as seen in the $35°$ rotation. The quasi invariant colour space performs best, as this approach takes only colour differences into account. This leads to a 95% repeatability rate at a $30°$ rotation and an 85% rate after the full $50°$ rotation.

4.2. Image Retrieval

For the retrieval experiment, the impact of the extraction of interest points in a retrieval scenario is examined. 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 18,000 images. Since the ALOI database contains images of objects 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.

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 five different retrieval tests is in the interest point extraction stage.

Image retrieval is performed by measuring the distance between the query image and every other image in the database. The difference between two images is determined by first calculating the Euclidean distances between each possible pair of (normalised) SIFT descriptors. The mean of the N smallest distances is taken to be the distance between the images (we use $N = 100$). The images are then ranked according to this distance. As the retrieval performance measure, the precision and recall values are calculated for subsets of retrieved images containing the $1, 2, \ldots, 30$ best matches to the query image.

Fig. 4 gives the precision recall graph of the overall retrieval rate. Colour based approaches outperform the illumination based methods, especially in low contrast, dark images, where illumination gets ambiguous and unstable. Beginning with the OCS colour space (OCS ScIv Harris), it can overcome this weakness resulting in more distinctive and more stable interest points. Using the HSI ScIv Harris, the colour only interest points outperform the OCS based ones, but lack distinctive interest points on uni-colour objects. Colour boosted OCS showed to perform best in the overall experiment. The prioritized salient colours and hence prioritized corners give more stable interest points than other approaches.

5. CONCLUSION

Using colour distances for corner measurement can shift the interest points to more 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.

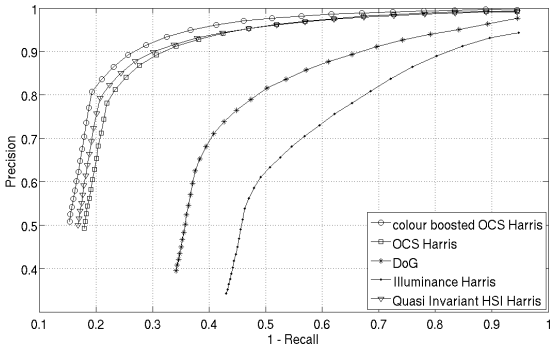


Fig. 4. Precision recall graph of the retrieval experiment.

6. REFERENCES

[1] C. Harris and M. Stephens, "A combined corner and edge detection," in *4th Alvey Vision Conference*, 1988, pp. 147–151.

[2] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval," *PAMI*, vol. 19, no. 5, pp. 530–535, 1997.

[3] T. Lindeberg, "Feature detection with automatic scale selection," *IJCV*, vol. 30, no. 2, pp. 79–116, 1998.

[4] K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors," *IJCV*, vol. 60, no. 1, pp. 63–86, 2004.

[5] Y. Dufournaud, C. Schmid, and R. Horaud, "Matching images with different resolutions," in *CVPR*, 2000, pp. 612–618.

[6] J. van de Weijer and T. Gevers, "Edge and corner detection by photometric quasi-invariants," *PAMI*, vol. 27, no. 4, pp. 625–630, 2005.

[7] P Montesinos, V Gouet, and R Deriche, "Differential invariants for color images," in *ICPR*, 1998, p. 838.

[8] C.Kenney, M. Zuliani, and B. Manjunath, "An axiomatic approach to corner detection," in *CVPR*, 2005.

[9] J. van de Weijer, T. Gevers, and A. Bagdanov, "Boosting color saliency in image feature detection," *PAMI*, vol. 28, no. 1, pp. 150–156, 2006.

[10] J.M. Geusebroek, G.J. Burghouts, and A.Smeulders, "The Amsterdam library of object images," *IJCV*, vol. 61, no. 1, pp. 103–112, 2005.

[11] K.Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *IJCV*, vol. 65, no. 1/2, pp. 43–72, 2005.