Lonely but Attractive: Sparse Color Salient Points for Object Retrieval and Categorization

Julian Stöttinger
Vienna University of Technology
Favoritenstr 9/183-2
Vienna, Austria
julian@prip.tuwien.ac.at

Allan Hanbury, Theo Gevers and Nicu Sebe
University of Amsterdam
Kruislaan 403
Amsterdam, The Netherlands
[hanbury|gevers|nicu]@science.uva.nl

Abstract

Local image descriptors computed in areas around salient points in images are essential for many algorithms in computer vision. Recent work suggests using as many salient points as possible. While sophisticated classifiers have been proposed to cope with the resulting large number of descriptors, processing this large amount of data is computationally costly.

In this paper, computational methods are proposed to compute salient points designed to allow a reduction in the number of salient points while maintaining state of the art performance in image retrieval and object recognition applications. To obtain a more sparse description, a color salient point and scale determination framework is proposed operating on color spaces that have useful perceptual and saliency properties. This allows for the necessary discriminative points to be located, allowing a significant reduction in the number of salient points and obtaining an invariant (repeatability) and discriminative (distinctiveness) image description.

Experimental results on large image datasets show that the proposed method obtains state of the art results with the number of salient points reduced by half. This reduction in the number of points allows subsequent operations, such as feature extraction and clustering, to run more efficiently. It is shown that the method provides less ambiguous features, a more compact description of visual data, and therefore a faster classification of visual data.

1. Introduction

Salient points, also referred to as *interest points*, are important in current solutions to computer vision challenges. In general, the current trend is toward increasing the number of points [26], applying several detectors or combining them [13, 19], or making the salient point distribution as dense as possible [16, 21]. While such a dense sampling has been shown to be effective in object recognition, these approaches basically shift the task of discarding the non-discriminative points to the classifier.

Dense sampling implies that a huge amount of data must be extracted from each image and processed. This is feasible when executed on a cluster of computers in a research environment. Nevertheless, there are environments in which the luxury of extensive computing power is not available. This is illustrated by the strong trend towards mobile computing on Netbooks, mobile phones and PDAs. Therefore, an important question is if it is possible to reduce the number of salient points extracted while still obtaining state of the art image retrieval or object recognition results.

Therefore, in this paper, computational methods are proposed to compute salient points, designed to allow a reduction in the number of salient points while maintaining state of the art performance in image retrieval and object recognition applications. The ability to choose the most discriminative points in an image is gained through including color information in the salient point determination process. To this end, a framework is presented for using color information to extract salient points and select a scale associated with each salient point. The color salient point and scale determination framework proposed operates on color spaces that have useful perceptual and saliency properties. This allows for the necessary discriminative points to be located, allowing a significant reduction in the number of salient points. In this way, an invariant (repeatability) and discriminative (distinctiveness) image description is obtained.

Experimental results are presented to demonstrate that fewer color salient points maintain state of the art performance for various applications. Firstly, we evaluate the repeatability of corner detection approaches using the evaluation framework of [14]. We show that the use of color increases the stability and distinctiveness of salient points in natural scenes under varying transformations so that fewer

978-1-4244-3993-5/09/$25.00 ©2009 IEEE

points maintain state of the art performance. We then show that by using fewer color salient points, we obtain improved retrieval of color images, by being more stable to lighting and shadowing effects. Finally, we evaluate object categorization using the PASCAL VOC dataset and show that the use of significantly fewer color salient points gives comparable performance to the best performing system in the challenges [26].

Section 2 gives an overview of the main approaches to extract salient points. Section 3 presents the theory behind the color salient point detection and scale calculation, with an overview of the color representations used in the experiments. With the experiments in Section 4, we demonstrate that significantly fewer color salient points perform equally well to existing state of the art techniques.

2. Related work

We give an overview of the successful approaches for detecting salient points based on intensity information and their extensions to use color information.

Blob detectors – based on the space-scale theory [25] and extended by [10] – rely on differential methods such as Laplacian of Gaussians (LoG), difference of Gaussians (DoG) and Determinant of Hessian (DoH) [10]. The result of blob detection using either LoG or DoG methods depends on the choice of a scale sampling rate which is analyzed in [11].

Maximally stable extremum regions (MSER) [12] are obtained by a watershed-like algorithm. Connected regions of a certain thresholded range are selected if they remain stable over a set of thresholds. The algorithm is efficient both in run-time performance and detection rate. The region saliency is measured as the number of thresholds where the region remains stable.

A widely used color based interest point detector is the Harris corner detector [7] which can be extended to RGB [15]. Instead of using the intensity gradient, the gradient for each RGB channel is determined. These values are summed and averaged using a Gaussian kernel with size σ_D. It is shown in [6] that it is the most stable interest point detector with respect to illumination changes, noise, rotation and viewpoint changes. It is successfully used in many applications including object tracking [5], visual information retrieval [18] and object-based queries. As suggested in [24], the second moment matrix can be computed using different color models. The first step is to determine the gradients of each RGB component, which are then transformed to the desired color space. This method forms the basis for the color salient point detector in Section 3.

Rugna et al. [18] suggest a method to extract scale-invariant interest points based on color information for texture classification. They build a color Gaussian pyramid using every channel separately. For every pyramid level and color channel, the Harris energy is calculated building a perceptual image of the texture. This method is independent of the color space used. The authors suggest the YUV or CIELAB color space.

Faille [4] proposes a shadow, shading, illumination color and specularities invariant interest point localization using the perceived RGB information as terms modeled as Lambertian and specular reflection, and expresses their geometric dependencies as a function of the light direction, surface normal and viewing direction. The approach is evaluated with the shadow-shading invariant HSI approach that we choose in Section 3.3. Unnikrishnan et al. [22] extract scale and illumination invariant blobs through color. The resulting regions are found by non-maxima suppression in the scale space pyramid.

3. Color Salient Point and Scale Detection

We first present the Harris corner detector generalized to color images, followed by the PCA method for determining the characteristic scale of the region surrounding an interest point in color images [20]. Finally we describe the two color spaces used for the experiments. In contrast to other color interest points used so far [1, 4, 5, 15, 18, 22] we are using an adapted color Harris detector in conjunction with an independent scale selection maintaining the main properties of the chosen color space [20]. Very relevant to our work is the research of van de Weijer et al. [24, 23]. They did preliminary work on incorporating color distinctiveness into the design of salient point detectors (see Section 3.3) and extended the RGB based Harris to be applied on arbitrary color spaces which is given in the next Section.

3.1. Color Salient Point Detector

The Harris corner detector introduced in [7] provides a cornerness measure for image data. It is based on the second moment matrix which describes the gradient distribution in the local neighborhood of a pixel in a greyscale image.

The second moment matrix can be computed in color spaces that are obtained by transformation from the RGB space [24]. The first step is to determine the gradients of each component of the RGB color system. This is done using a convolution with a differentiation kernel of size σ_D. The gradients are then transformed into the desired color 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 σ_I. To achieve scale invariance, scale normalization is done using the factor σ_D.

In symbolic form, an arbitrary color space C is used with its n components $[c_1, \ldots, c_n]^T$. The second moment matrix

M computed at position $\mathbf{x}$ in a color image is then

$$M(\mathbf{x}, \sigma_I, \sigma_D) = \sigma_D^2 G(\sigma_I) \otimes \begin{bmatrix} L_x^2(\mathbf{x}, \sigma_D) & L_x L_y(\mathbf{x}, \sigma_D) \\ L_x L_y(\mathbf{x}, \sigma_D) & L_y^2(\mathbf{x}, \sigma_D) \end{bmatrix} \tag{1}$$

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

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

where $c_{i,x}$ and $c_{i,y}$ denote the components of the transformed color channel gradients, and where the subscript x or y indicates the direction of the gradient. As shown in several experiments [14], the relation $3\sigma_D = \sigma_I$ performs best. Given a second moment matrix $M(\mathbf{x}, \sigma_I, \sigma_D)$, the Harris energy is calculated based on the trace and determinant of this matrix:

$$C_H(\mathbf{x}, \sigma_I, \sigma_D) = det(M) - \alpha \cdot trace^2(M) \tag{3}$$

where the matrices M are calculated using the parameters specified as arguments of the C_H function. The constant α indicates the slope of the *zero line*, i.e. the border between corner and edge. We show in Section 4 that this measure gives a powerful saliency indication for the choice of important salient points. In the next section, we suggest a way to estimate the characteristic scale of a salient point based on this measurement.

3.2. Color Scale Decision

Mikolajczyk and Schmid [14] applied the Laplacian of Gaussian for automatic scale selection. This is referred to as the *Harris Laplacian* detector. The scale space of the Harris function is built by iteratively calculating the Harris energy under varying σ_D and σ_I.

We apply *principal component analysis* (PCA) to reduce the transformed color image I to a single channel finding the most representing projection vector and reducing the projective representation error $\widehat{I}(\mathbf{x}) = \nu_\lambda I(\mathbf{x})^T$ [2]. In this scope the projection aims to maintain both salient colors and relative color differences. Therefore we make use of the "saliency implies rarity" principle from [8].

The Laplacian of Gaussian function Λ has been used to detect the characteristic scale automatically [10]. Λ is defined by

$$\Lambda(\mathbf{x}, \sigma_D) = \left(\frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \right) \otimes G(\sigma_D) \otimes \Gamma(\sigma_D) \tag{4}$$

where $\Gamma(\sigma_D)$ is the circularly symmetric raised cosine kernel, which is defined for each location x_i, y_i in the patch with the center x_c, y_c and the radius σ_D

$$\Gamma(\sigma_D) = \frac{1 + (cos(\frac{\pi}{\sigma_D} \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2})}{3} \tag{5}$$

A convolution with this kernel gives smoother borders than the Gaussian kernel G for scale decision [9].

A characteristic scale of a possible region is found if both the Harris energy and the Laplacian of Gaussian are extrema [20]. Aiming for just one region per location covering the characteristic structure around, we use the following decision criteria:

$$\widehat{R}(\mathbf{x}) = \begin{pmatrix} \max_s \left[\widehat{E}(\mathbf{x}, s) \right] \\ 3t^{\arg\max_s \left[(\widehat{\Lambda}(\mathbf{x}, s)) \right]} \sigma_D \end{pmatrix} \tag{6}$$

where, having chosen constants σ_I, σ_D and t, the functions are $\widehat{E}(\mathbf{x}, s) = C_H(\mathbf{x}, t^s\sigma_I, t^s\sigma_D)$ and $\widehat{\Lambda}(\mathbf{x}, s) = \Lambda(\mathbf{x}, t^s\sigma_D)$. The resulting vector function $\widehat{R}(\mathbf{x})$ defines all candidates for salient points and the corresponding region size. The Harris energy is the measure used to characterize salient points, and is used as a decision criterion for discarding less salient points. The characteristic scale is estimated independently of the scale in which the highest Harris energy occurs.

3.3. Color Spaces

We calculate the proposed salient points in two color spaces encoding luminance and chroma information separately. We use the color space HSI proposed in [24], derived from the Opponent Color Space (OCS) defined as

$$OCS = \begin{pmatrix} o_1 \\ o_2 \\ o_3 \end{pmatrix} = \begin{pmatrix} \frac{R-G}{\sqrt{2}} \\ \frac{R+G-2B}{\sqrt{6}} \\ \frac{R+G+B}{\sqrt{3}} \end{pmatrix} \tag{7}$$

This orthonormal transformation into OCS provides specular variance.

A polar transformation on o_1 and o_2 of the OCS leads to the HSI color space

$$HSI = \begin{pmatrix} h \\ s \\ i \end{pmatrix} = \begin{pmatrix} \tan^{-1}\left(\frac{o_1}{o_2}\right) \\ \sqrt{o_1^2 + o_2^2} \\ o_3 \end{pmatrix} \tag{8}$$

The derivative of the hue component h is both the shading and the specular quasi-invariant [24], as it is both perpendicular to the shadow-shading direction and the specular direction.

As proposed in [23], colors have different occurrence probabilities $p(v)$ and therefore, different information content $I(v)$ of the descriptor v:

$$I(v) = -log(p(v)) \tag{9}$$

The idea behind *color boosting* is to boost rare colors to have a higher saliency in the cornerness measurement. Looking for rare colors, statistics for the Corel Database containing 40000 color images showed that the three dimensional color distribution was remarkably significant. The color boosting transformation is chosen that vectors of equal saliency lead to vectors of equal length. A boosting function can be found so that pixel information in the image data has equal impact on the saliency function as its information content. The strength of gradients is considered as the decorrelated information content [24].

These color spaces provide either shading and specular quasi-invariance or a occurrence probability based saliency measure. Therefore we use these two color spaces to extract the color salient points used in the experiments in the following Section.

4. Experiments

The experiments aim to demonstrate that state of the art results can be obtained when using significantly fewer color salient points. We first demonstrate the stability of these color salient points by carrying out extensive repeatability experiments on natural scenes in Section 4.1. We show that color information increases the stability of salient points especially for lighting changes. In an image retrieval experiment in Section 4.2, we show that fewer color salient points outperform intensity-based Harris Laplacian salient points. In addition, the increased ability of color points to retrieve colorful objects under lighting changes is demonstrated. Going into "real world" experiments with natural images, we challenge the color salient points in a large public object categorization challenge. To show the impact of the location of detectors in different applications, we do not use color based description but use illumination based SIFT and SPIN descriptors only.

In the experiments, we denote the salient points based on the HSI quasi invariant coordinates as *light invariant points* and the salient points based on the color boosted OCS coordinates as *color boosted points*. When we refer to both of them, we refer to them as *color points*. In Section 4.1, we add RGB points for comparison. As the state of the art reference, we use the *Harris Laplacian*. For the Harris Laplacian we use the latest implementation[1] used and evaluated in [14]. All experiments are carried out with the same parameters $\sigma_D = 1$, $s = 10$, $t = \sqrt{2}$ as suggested in the literature and defined in Section 3. In case we choose a subset of the provided points, we order the points by their Harris energy (see Eq. 3). We show the color based Harris energy gives a more reliable decision criteria for reducing features than the illumination based counterpart does.

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

4.1. Repeatability

Mikolajczyk and Schmid [14] suggest a test for the quality of salient points: they measure the repeatability of salient points under different challenges. A challenge consists of a set of images, where one is the reference image, and the other images show the same scene under predefined changes, including blur, rotation, zoom, viewpoint change, jpeg compression and lighting[2]. An example set challenging viewpoint change is shown in Figure 1. The repeatability rate is defined as the ratio between the number of actual corresponding salient points and the total number of salient points that occur in the area common to both images. More salient points give a higher probability of overlaps and tend to have a higher repeatability rate. In Fig. 2 the averaged results of the experiments with the suggested test sets are shown. We carry out the repeatability challenges on a varying number of features. From 100 points on, the results become stable and go up to 6000 points, when statistically all the pixels are covered by at least 10 salient points at once. We denote this as a *dense distribution* of salient points. We evaluate the approaches for selecting salient points. For the Harris Laplacian, literature suggests a fixed threshold on the Harris energy. This leads to a variable number of points for the images of the dataset based on their contrast. This increases stability for certain challenges, but is a drawback for others e.g. varying contrast which happens e.g. at lighting challenges. We see that 1000 color points reach Harris Laplacian performance with the suggested parameters (mean number of points: 2688 ([763,9191] $\pm$ 2536) and even outperform its dense distribution of 6000 points per image. Comparing light invariant points with Harris Laplacian, 100 light invariant points are enough to outperform the state of the art.

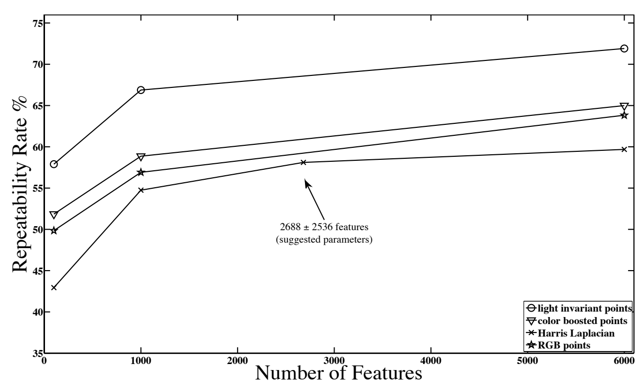

Figure 2. The mean repeatability rate of the 18 repeatability challenges per number of points.

Figure 3 shows the mean repeatability over the five datasets with varying lighting. Increasing the number of

[2]http://lear.inrialpes.fr/people/mikolajczyk/Database

Figure 1. “Graffiti” test set used in the repeatability experiment challenging change of viewpoint.

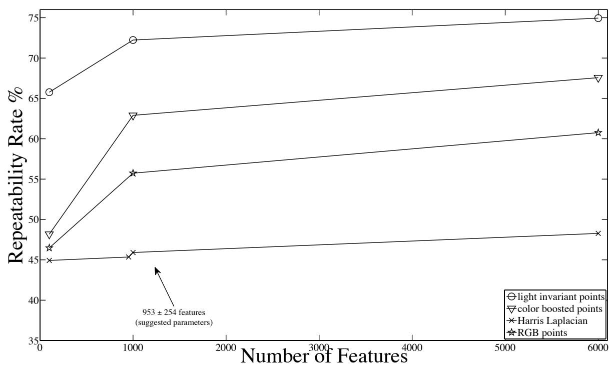

Figure 3. The mean repeatability rate of the 5 data-sets challenging lighting changes only.

Harris Laplacian points does not improve the repeatability against light changes significantly. In contrast, light invariant points remain more stable for all chosen number of points. Generally, color boosted points prove to be less repeatable than the HSI points, which is reasonable as their saliency function is variant with respect to illumination changes and focuses on the occurrence probability. The RGB points have the lowest repeatability among the colour interest points tested, and are therefore omitted from the subsequent experiments.

These results show that the Harris energy of the color points gives a better saliency measurement for reducing features. As already stated, increasing the number of color points increases the repeatability for the color points, but even with a very sparse description, we have reasonable results outperforming the Harris Laplacian. A complete description is not necessarily a matter of quantity but of the reasonable distribution of the points.

4.2. Image Retrieval

This experiment evaluates the impact of different color spaces in retrieval scenarios with varying illumination direction and intensity. We use the Amsterdam Library of Object Images (ALOI)[3] which provides images of 1000 objects under supervised, predefined conditions on a dark background yielding a total of 110250 images for the collection.

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

(a) ALOI object 225 (b) ALOI object 245 (c) ALOI object 584

Figure 4. Sparse color points retrieve object 225 (a) perfectly with rank 8-13 going to object 245 (b). Dense points perform worse shifting object 584 (c) to rank 2-8.

Example images are shown in Figure 5. We use the part of the dataset that provides images under eight predefined illumination conditions for each object, where illumination direction and illumination intensity is varied. With these illumination changes, intensity based approaches suffer from instability and many ambiguous descriptions of shadowing effects. This experiment is carried out with 7000 images as ground truth set and 1000 query images, having thus seven true positives for every object class and query image.

The salient point approaches evaluated provide the locations and scales for the subsequent calculation of SIFT descriptors. For the matching, the similarity between two images is determined by a simple *knn* algorithm ranking the nearest 30 Euclidean weighted distances of 100 best matches. The only difference between the evaluated approaches is in the stage of salient point extraction.

We show that there is a certain minimum number of features that is necessary to discriminate an object from 999 other objects. More importantly we see that too many features make the description ambiguous. To discover this crucial global parameter, we do the retrieval experiment while varying the maximum number of salient points to be extracted, where a maximum number of N salient points implies that the N salient points with the largest Harris energies are extracted. If fewer than N salient points are detected for an image, then all are used. We begin with all extractable salient points (all of the up to 22117 maxima of the Harris energy per image) and then use $N = 1000, 500, 200, 100, 50, 10$. We retrieve the most similar images, considering them as ranked results for the retrieval evaluation. We consider the precision and recall for the top 30 retrieved images and show the mean of the resulting F1 score plotted against the number of salient points in Figure 6. We see that every approach has its best performing

Figure 5. Example images from the ALOI database.

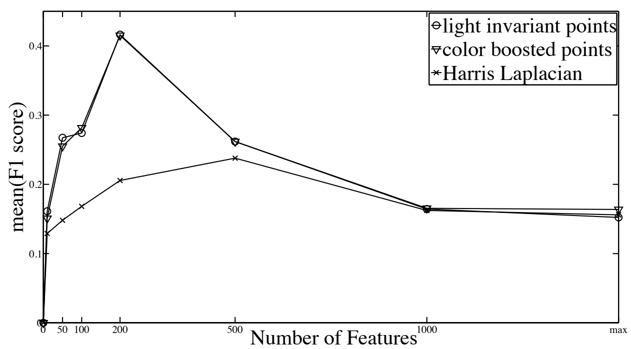

Figure 6. Mean F1 score of 30 ranks under different maximum number of features.

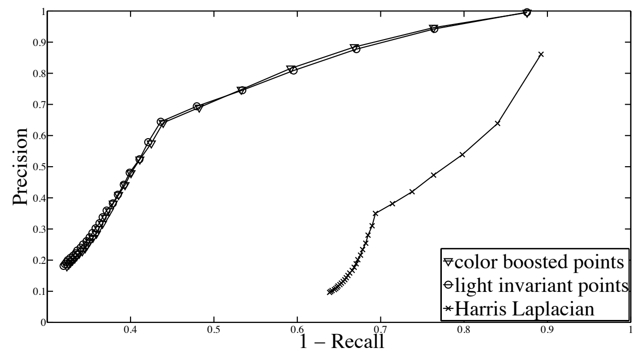

Figure 7. Precision and recall of retrieval performance for best performing parameters of the color points and the suggested parameters of the Harris Laplacian.

	avg. precision	avg. nr. points
Harris Laplacian	0.52	381
color boosted points	0.82	192
HSI points	0.82	193

Table 1. Average of number of points extracted and the average precision in the image retrieval experiment.

number of salient points. The important point is that after a maximum, the performance decreases with increasing number of points.

Figure 4 shows a specific example of this decrease in performance with an increasing number of salient points. Object 225 is shown in Fig. 4(a): It is retrieved perfectly with the first 7 ranks being correct for 200 light invariant points. The next candidate with the 2nd best ranks is object 245 (Figure4(b)) for this set of parameters. This is intuitively correct as it contains similar texture. With 200 light invariant points, object 225 does not appear in ranks 1–7 for queries using any of the other 999 objects. Taking all the 8775 features available, object 225 appears in 43 queries in the top 7 ranks, worsening the result significantly. For the query by object 225 itself, it still ranks one right candidate at the first rank, having the following 7 from object 584 (see Fig. 4(c)). As the only distinct features, the spikes at the border of object 225 and on the head of object 584 remain. The other features become more ambiguous the more points we consider. It is clear from Fig. 6 that a higher performance is achieved for a lower number of color salient points than for the Harris Laplacian points. The precision-recall graph showing averages over 1000 queries for the best performing parameters of the color points and the suggested parameters of the Harris Laplacian (max. of 200 color points, 381 [12,8873] $\pm$ 393 Harris Laplacian points) is given in Figure 7, with the average precision shown in Table 1. On average, half of the number of color points are used to almost solve this retrieval scenario perfectly. Compared to the Harris Laplacian based approach, reducing the number of points to a half reduces the computational complexity significantly.

4.3. Object categorization

One of the most successful approaches to object categorization is the bag of keypoints in combination with SVM classifiers [26] — the best performing methods at the PASCAL Visual Object Classes Challenge 2006 [3] and later used variations on this approach.

We evaluate the color salient points on the dataset from the PASCAL VOC 2007 challenge[4]. This dataset consists of 9963 images, where 5011 images form the annotated training set. The test set contains 4952 images which are used to evaluate the performance of the framework. The number of objects in one image is not fixed, the whole dataset contains 12608 objects. Twenty classes of object are annotated with ground truth. Example images are shown in Figure 9. As a benchmark we use the algorithms that are evaluated in more detail by Zhang et al. [26]. They use a Harris Laplacian detector, a combination of SIFT and SPIN descriptors using a bags of keypoints approach and an EMD Kernel SVM for classification. The workflow of the algorithm is shown in Figure 8. We use the best performing parameters

[4] http://www.pascal-network.org/challenges/VOC/voc2007/

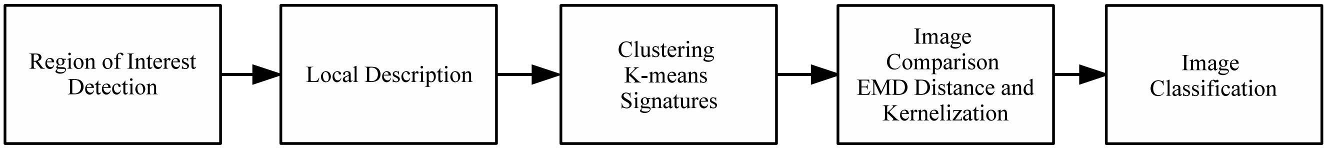

Figure 8. Flowchart of the approach of Zhang et al. [26] used in the object categorization experiment.

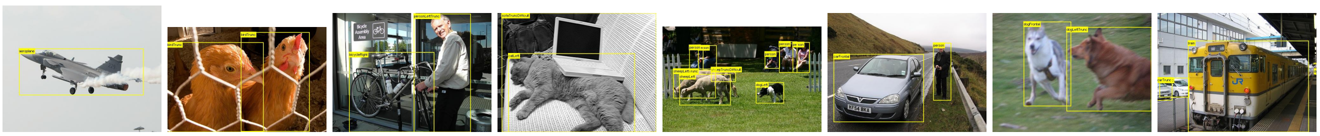

Figure 9. Annotated sample images from the VOC 2007 dataset.

of their evaluation. Image signatures consisting of 40 clusters of these descriptors are extracted. Clustering is done using the k-means algorithm. The earth mover's distance (EMD) [17] showed to perform best for the task of estimating the similarity between image signatures. These image distances are incorporated into a one-against-all SVM classifier. We wish to test the effect of using salient points obtained from luminance and color information on the categorization and calculation time performance. Only the first step in the flowchart in Figure 8 is changed, all succeeding steps of the approach are carried out identically. An example showing the color points and Harris Laplacian of an image from the VOC 2007 dataset is given in Figure 10. For this image, the color salient point detectors focus on the more colorful foreground objects. Fewer salient points are found in the background by the color salient point detectors than by the Harris Laplacian detector.

We test the quality of each single one-against-all classifier by carrying out 10 fold cross validation on the VOC 2007 training set. The second column of Table 2 shows the discrimination accuracy for different numbers and types of salient points and descriptors, averaged over 20 one-against-all classifiers. No matter which description we feed into the SVM, the classifier manages to reach about 93% accuracy on the 2 class problem. The accuracy when categorizing the test data into one of 20 classes is shown in the third column of Table 2. For this experiment, reducing the number of Harris Laplacian points by about 50% gives around 60% of the original categorization performance. This does not hold for color salient points: we keep the 400 salient points with the highest Harris energy per image and can maintain the performance of the richer description (800 points).

Therefore we argue that the color points are more distinct and discriminative, even when intensity based descriptors are used. It is shown that the use of color in the detection phase does not degrade the categorization ability of the one-against-all SVM classifier and the description is as complete as for the best performing Harris Laplacian detector. The classifier is able to discriminate between the given

SIFT	discrimination	categorization	number of points
Harris Laplacian	93.12 $\pm$ 2.52%	79.5 $\pm$ 15.5%	771 $\pm$ 531
	92.41 $\pm$ 2.65%	54.6 $\pm$ 20.7%	387 $\pm$ 72
light invariant points	93.27 $\pm$ 2.17%	81.7 $\pm$ 10.6%	800
	93.54 $\pm$ 2.34%	80.9 $\pm$ 11.4%	400
color boosted points	93.49 $\pm$ 2.28%	83.0 $\pm$ 10.1%	800
	93.41 $\pm$ 2.44%	**83.1 $\pm$ 10.4%**	400
SPIN			
Harris Laplacian	92.95 $\pm$ 2.64%	66.2 $\pm$ 17.8%	771 $\pm$ 531
	92.19 $\pm$ 2.8%	38.4 $\pm$ 12.9%	387 $\pm$ 72
light invariant points	**93.16 $\pm$ 2.61%**	**68.9 $\pm$ 18.3%**	800
	93.08 $\pm$ 2.56%	68.3 $\pm$ 18.9%	400
color boosted points	93.13 $\pm$ 2.71%	68.8 $\pm$ 17.2%	800
	93.04 $\pm$ 2.62%	68.7 $\pm$ 16.4%	400
SIFT + SPIN			
Harris Laplacian	93.50 $\pm$ 2.4%	85.9$\pm$ 9.9%	771 $\pm$ 531
	92.83 $\pm$ 2.75%	54,8 $\pm$ 18.5%	387 $\pm$ 72
light invariant points	93.52 $\pm$ 2.61%	**86.6 $\pm$ 8.7%**	800
	93.57 $\pm$ 2.37%	86.5 $\pm$ 8.5%	400
color boosted points	93.49 $\pm$ 2.65%	86.2 $\pm$ 8.9%	800
	93.47 $\pm$ 2.38%	86.4$\pm$ 8.4%	400

Table 2. Discrimination accuracy of the classifier and the categorization accuracy of the challenge as average $\pm$ standard deviation over classes.

object classes equally well, while training on significantly fewer descriptors. The number of salient points are an indication for the runtime of the system. Every step of the object categorization (see Fig. 8) has to deal only with about half of the data as the state of the art does, which diminishes the runtime significantly.

5. Conclusion

In this paper, computational methods have been introduced to allow the usage of fewer but more distinctive salient points for object retrieval and categorization. These distinctive points are obtained by making use of color information.

Extensive experimental results show that a sparser but equally informative representation, obtained by making use of color information, can be directly passed to current and successful image retrieval and object categorization frameworks, which then obtain state of the art results while processing significantly less data. When using color salient point detectors for object categorization, the same performance is obtained using about half the number of color in-

(a) original (b) light invariant points (c) color boosted points (d) Harris Laplacian

Figure 10. VOC 2007 image number 5221 and the salient points used in the object categorization experiment.

terest points compared to greyscale interest points. Such a reduction in the amount of data to be processed is useful in applications for which limited computing power is available.

Acknowledgments

This work was partly supported by the Austrian Research Promotion Agency (FFG) project 815994, and the CogVis[5] Ltd.

References

[1] A. E. Abdel-Hakim and A. A. Farag. Csift: A sift descriptor with color invariant characteristics. In *CVPR*, pages 1978–1983, 2006. 2

[2] K. Delaca, M. Grgic, and S. Grgic. Independent comparative study of PCA, ICA, and LDA on the FERET data set. *IJIST*, 15:252–260, 2005. 3

[3] M. Everingham, A. Zisserman, C. K. I. Williams, and L. Van Gool. The PASCAL Visual Object Classes Challenge 2006 (VOC2006) Results. 6

[4] F. Faille. Stable interest point detection under illumination changes using colour invariants. In *BMVC*, 2005. 2

[5] P. Gabriel, J.-B. Hayet, J. Piater, and J. Verly. Object tracking using color interest points. In *AVSS*, pages 159–164, 2005. 2

[6] V. Gouet and N. Boujemaa. Object-based queries using color points of interest. In *CVPRW*, pages 30 – 36, 2001. 2

[7] C. Harris and M. Stephens. A combined corner and edge detection. In *AVC*, pages 147–151, 1988. 2

[8] T. Kadir and M. Brady. Saliency, scale and image description. *IJCV*, 45(2):83–105, 2001. 3

[9] C. Kenney, M. Zuliani, and B. Manjunath. An axiomatic approach to corner detection. In *CVPR*, pages 191–197, 2005. 3

[10] T. Lindeberg. Feature detection with automatic scale selection. *IJCV*, 30(2):79–116, 1998. 2, 3

[11] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004. 2

[12] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. In *BMVC*, pages 384–393, 2002. 2

[13] K. Mikolajczyk, B. Leibe, and B. Schiele. Multiple object class detection with a generative model. In *CVPR*, pages 26–36, 2006. 1

[14] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool. A comparison of affine region detectors. *IJCV*, 65(1/2):43–72, 2005. 1, 3, 4

[15] P. Montesinos, V. Gouet, and R. Deriche. Differential invariants for color images. In *ICPR*, page 838, 1998. 2

[16] E. Nowak, F. Jurie, and B. Triggs. Sampling strategies for bag-of-features image classification. In *ECCV 2006*, pages 490–503, 2006. 1

[17] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *IJCV*, 40(2):99–121, 2000. 7

[18] J. D. Rugna and H. Konik. Color interest points detector for visual information retrieval. In *SPIE*, volume 4672, pages 144–151, 2002. 2

[19] J. Sivic, B. Russell, A. A. Efros, A. Zisserman, and B. Freeman. Discovering objects and their location in images. In *ICCV*, pages 370–377, 2005. 1

[20] J. Stöttinger, A. Hanbury, N. Sebe, and T. Gevers. Do colour interest points improve image retrieval? In *ICIP*, pages 169–172, 2007. 2, 3

[21] T. Tuytelaars and C. Schmid. Vector quantizing feature space with a regular lattice. In *ICCV*, 2007. 1

[22] R. Unnikrishnan and M. Hebert. Extracting scale and illuminant invariant regions through color. In *BMVC*, September 2006. 2

[23] J. van de Weijer, T. Gevers, and A. Bagdanov. Boosting color saliency in image feature detection. *PAMI*, 28(1):150–156, 2006. 2, 3

[24] J. van de Weijer, T. Gevers, and J. M. Geusebroek. Edge and corner detection by photometric quasi-invariants. *PAMI*, 27(4):625–630, 2005. 2, 3, 4

[25] A. P. Witkin. Scale-space filtering. In *IJCAI*, pages 1019–1022, 1983. 2

[26] J. Zhang, M. Marszałek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories: A comprehensive study. *IJCV*, 73(2):213–238, 2007. 1, 2, 6, 7

[5] http://www.cogvis.at/