

COMPARING SALIENT POINT DETECTORS

Nicu Sebe and Michael S. Lew

Leiden Institute of Advanced Computer Science
NielsBohrweg 1, 2333 CA, Leiden, The Netherlands
{nicu mlew}@liacs.nl

ABSTRACT

The use of salient points in content-based retrieval allows an image index to represent local properties of the image. Classic corner detectors can also be used for this purpose but they have drawbacks when are applied to various natural images mainly because visual features do not need to be corners and corners may gather in small regions. In this paper, we present a salient point detector using wavelet transform and we compare it with two corner detectors using two criteria: repeatability rate and information content. We determine which detector gives the best results and show that it satisfies the criteria well.

1. INTRODUCTION

Many computer vision tasks rely on low level features. A wide variety of feature detectors exist and results can vary enormously depending on the detector used. An image is "summarized" by a set of features, the image index, to allow fast querying. Local features are of interest since they lead to an index based on local properties of the image. The feature extraction is limited to a subset of the image pixels, the interest points, where the image information is supposed to be the most important [7, 9]. Besides saving time in the indexing process, these points may lead to a more discriminant index because they are related to the visually most important parts of the image.

Schmid and Mohr [7] introduced the notion of interest point in image retrieval. In order to detect such points they use the Harris corner detector [3]. The basic idea is to use the auto-correlation function in order to determine locations where the signal changes in two directions. A matrix related to the auto-correlation function which takes into account first derivatives of the signal on a window is computed. The eigenvectors of this matrix are the principal curvatures of the auto-correlation function. Two significant values indicate the presence of an interest point.

Different interest point detectors are evaluated and compared in [8]. Besides the Harris corner detector and an improved variant of it the authors also consider the detectors proposed by Heitger [4], Förstner [2], and Horaud [5]. The authors [8] concluded that the best results are provided by the Harris detector and therefore we use it for our benchmarking.

Corner detectors are designed for robotics and shape recognition and they have drawbacks when are applied to natural images. Corners may gather in small regions and visual features do not need to be corners. For these reasons, corner points may not represent the most interesting subset of pixels. Salient points should be related to any visual interesting part of the image whether it is smoothed or corner-like. Moreover, to describe different parts of the image the set of salient points should not be clustered in few regions. In our approach, we consider wavelet representations which express image variations at different resolutions.

In order to evaluate the results of different interest point detectors two criteria are considered: repeatability rate and information content. Repeatability rate evaluates the geometric stability of points under different image transformation. Information content measures the distinctiveness of greylevel pattern at an interest point. A local pattern is described using rotationally invariant combinations of derivatives. The entropy of these invariants is measured for a set of interest points.

2. WAVELET-BASED SALIENT POINTS

The wavelet representation gives information about the variations in the image at different scales. We want to extract salient points from any part of the image where something happens at any resolution. A high wavelet coefficient at a coarse resolution corresponds to a region with high global variations. The idea is to find a relevant point to represent this global variation by looking at wavelet coefficients at finer resolutions.

A wavelet is an oscillating and attenuated function with zero integral. We study the image I at the scales (or resolutions) $1/2, 1/4, \dots, 2^j, j \in \mathbb{Z}$ and $j \leq -1$. The wavelet detail image $W_{2^j}I$ is obtained as the convolution of the image with the wavelet function dilated at different scales. We considered orthogonal wavelets with compact support. First, this assures that we have a complete and non-redundant representation of the image. Second, we know from which signal points each wavelet coefficient at the scale 2^j was computed. We can further study the wavelet coefficients for the same points at the finer scale 2^{j+1} . There is a set of coefficients at the scale 2^{j+1} com-

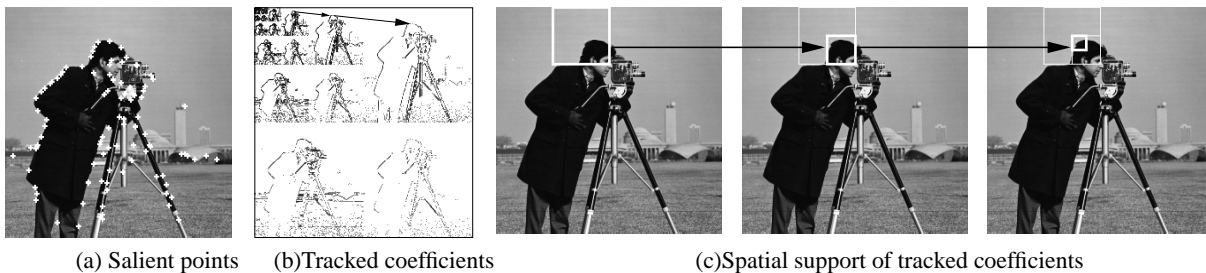


Figure 1: Salient points extraction

puted with the same points as a coefficient $W_{2^j}I(n)$ at the scale 2^j . We call this set of coefficients the children of the coefficient $W_{2^j}I(n)$: $C(W_{2^j}I(n)) = \{W_{2^{j+1}}I(k), 2n \leq k \leq 2n + 2p - 1\}$ where p is the wavelet regularity and $0 \leq n < 2^j N$ with N the length of the signal.

Each wavelet coefficient $W_{2^j}I(n)$ is computed with $2^{-j}p$ signal points. It represents their variation at the scale 2^j . Its children coefficients give the variations of some particular subsets of these points (with the number of subsets depending on the wavelet). The most salient subset is the one with the highest wavelet coefficient at the scale 2^{j+1} , that is the maximum in absolute value of $C(W_{2^j}I(n))$. In our salient point extraction algorithm, we consider this maximum, and look at his highest child. Applying recursively this process, we select a coefficient $W_{2^{-1}}I(n)$ at the finer resolution $1/2$ (Figure 1 (b) and (c)). Hence, this coefficient represents $2p$ signal points. To select a salient point from this tracking, we choose among these $2p$ points the one with the highest gradient. We set its saliency value as the sum of the absolute value of the wavelet coefficients in the track:

$$saliency = \sum_{k=1}^{-j} |C^{(k)}(W_{2^j}I(n))|, -\log_2 N \leq j \leq -1 \quad (1)$$

The tracked point and its saliency value are computed for every wavelet coefficient. A point related to a global variation has a high saliency value, since the coarse wavelet coefficients contribute to it. A finer variation also leads to an extracted point, but with a lower saliency value. We then need to threshold the saliency value, in relation to the desired number of salient points. We first obtain the points related to global variations; local variations also appear if enough salient points are requested.

The salient points extracted by this process depend on the wavelet we use. Haar is the simplest wavelet function, so it is the fastest for execution. However, some localization drawbacks can appear with Haar due to its non-overlapping wavelets at a given scale. This can be avoided with the simplest overlapping wavelet, Daubechies4 [1].

In Figure 1 (a) we present the salient points using the Haar transform. Note that our method extracts salient points not only in the foreground but also in the background where some smooth details are present.

3. REPEATABILITY

Repeatability is defined by the image geometry. Given a 3D point P and two projection matrices M_1 and M_2 , the projections of P into two images I_1 and I_2 are $p_1 = M_1P$ and $p_2 = M_2P$. The point p_1 detected in image I_1 is repeated in image I_2 if the corresponding point p_2 is detected in image I_2 . To measure the repeatability, a unique relation between p_1 and p_2 has to be established. In the case of a planar scene this relation is defined by an homography: $p_2 = H_{21}p_1$.

The percentage of detected points which are repeated is the *repeatability rate*. A repeated point is not in general detected exactly at position p_2 , but rather in some neighborhood of p_2 . The size of this neighborhood is denoted by ε and repeatability within this neighborhood is called ε -*repeatability*. Moreover, to measure the number of repeated points, we have to take into account that the observed scene parts differ in the presence of changed imaging conditions, such as image rotation or scale change. The salient points which cannot be observed in both images corrupt the repeatability measure and therefore, only the points which are detected in the common scene part should be used to compute the repeatability. Points d_1 and d_2 which are detected in the common part of images I_1 and I_2 are defined by $\{d_1\} = \{p_1 | H_{21}p_1 \in I_2\}$ and $\{d_2\} = \{p_2 | H_{12}p_2 \in I_1\}$, where $\{p_1\}$ and $\{p_2\}$ are the points detected in images I_1 and I_2 , respectively. The set of point pairs (d_2, d_1) which correspond within an ε -neighborhood is $D(\varepsilon) = \{(d_2, d_1) | dist(d_2, H_{21}d_1) < \varepsilon\}$.

The number of detected points may be different for the two images. For example, in the case of a scale change more salient points are detected on the high resolution image. Only the minimum number of salient points (the number of salient points of the coarse image) can be repeated. In this conditions, the repeatability rate for image I_2 is given by:

$$r(\varepsilon) = \frac{|D(\varepsilon)|}{\min(|\{d_1\}|, |\{d_2\}|)} \quad (2)$$

where $|\{d_1\}|$ and $|\{d_2\}|$ are the numbers of points detected in the common part of images I_1 and I_2 , respectively. One can easily verify that $0 \leq r(\varepsilon) \leq 1$.

4. INFORMATION CONTENT

Information content is a measure of the distinctiveness of a salient point. Distinctiveness is based on the likelihood of a greyvalue descriptor computed at the point within the population of all observed salient point descriptors. Given one image, a descriptor is computed for each of the detected salient points and the information content will measure the distribution of these descriptors. If all descriptors are spread out, information content is high and matching is likely to succeed. On the other hand, if all descriptors are close to each other, the information content is low and matching can easily fail as any point can be matched to any other.

Information content of the descriptors is measured using entropy. The more spread out the descriptors are, the larger the entropy is. Entropy measures average information content. In information theory the information content of a message i is inversely related to its probability and is defined as $I = -\log(p_i)$. The average information content per message of a set of messages is $-\sum_i p_i \log(p_i)$ which is the entropy.

In the case of salient points we would like to know how much average information content a salient point "has" as measured by its greylevel pattern. The more distinctive the greylevel patterns are, the larger the entropy is. To measure the distribution of local greyvalue patterns at salient points, we have to describe a measure which describes such a pattern. In order to have rotation invariant descriptors, we chose to characterize salient points by local greyvalue rotation invariants which are combinations of derivatives. We computed the "local jet" [6] which is consisted of the set of derivatives up to N th order. These derivatives describe the intensity function locally and are computed stably by convolution with Gaussian derivatives. The local jet of order N at a point $\mathbf{x} = (x, y)$ for an image I and a scale σ is defined by: $J^N[I](\mathbf{x}, \sigma) = \{L_{i_1 \dots i_n}(\mathbf{x}, \sigma) | (\mathbf{x}, \sigma) \in I \times R^+\}$, where $L_{i_1 \dots i_n}(\mathbf{x}, \sigma)$ is the convolution of image I with the Gaussian derivatives $G_{i_1 \dots i_n}(\mathbf{x}, \sigma)$, $i_k \in \{x, y\}$ and $n = 0, \dots, N$.

In order to obtain invariance under the group $SO(2)$ (2D image rotation), Koenderink and van Doorn [6] and ter Haar Romeny, et al., [10] compute differential invariants from the local jet:

$$\vec{v}[0 \dots 3] = \begin{bmatrix} L_x L_x + L_y L_y \\ L_{xx} L_x L_x + 2L_{xy} L_x L_y + L_{yy} L_y L_y \\ L_{xx} + L_{yy} \\ L_{xx} L_{xx} + 2L_{xy} L_{xy} + 2L_{yy} L_{yy} \end{bmatrix} \quad (3)$$

The computation of entropy requires a partitioning of the space of \vec{v} . Partitioning is dependent on the distance measure between descriptors and we consider the approach described by Schmid, et al. [8]. The distance we used is the Mahalanobis distance given by: $d_M(\vec{v}_1, \vec{v}_2) =$

$\sqrt{(\vec{v}_1 - \vec{v}_2)^T \Lambda^{-1} (\vec{v}_1 - \vec{v}_2)}$, where \vec{v}_1 and \vec{v}_2 are two descriptors and Λ is the covariance of \vec{v} . The covariance matrix Λ is symmetric and positive definite. Its inverse can be decomposed into $\Lambda^{-1} = P^T D P$ where D is diagonal and P an orthogonal matrix. Furthermore, we can define the square root of Λ^{-1} as $\Lambda^{-1/2} = D^{1/2} P$ where $D^{1/2}$ is a diagonal matrix whose coefficients are the square roots of the coefficients of D . The Mahalanobis distance can then be rewritten as: $d_M(\vec{v}_1, \vec{v}_2) = \|D^{1/2} P(\vec{v}_1 - \vec{v}_2)\|$. The distance d_M is the norm of difference of the normalized vectors: $\vec{v}_{norm} = D^{1/2} P \vec{v}$. This normalization allows us to use equally sized cells in all dimensions. This is important since the entropy is directly dependent on the partition used. The probability of each cell of this partition is used to compute the entropy of a set of vectors \vec{v} .

5. RESULTS

In our experiments we considered 4 salient point detectors. In Section 2 we introduced two salient point detectors using wavelets: Haar and Daubechies4. For benchmarking purposes we also considered the Harris corner detector [3] and a variant of it called PreciseHarris, introduced by Schmid, et al. [8]. The difference between the last two detectors is given by the way the derivatives are computed. Harris computes derivatives by convolving the image with the mask $[-2 \ -1 \ 0 \ 1 \ 2]$ whereas PreciseHarris uses derivatives of the Gaussian function instead.

We used a set of 1000 images taken from the Corel database. We used this database because it represents a widely used set of photos by both amateur and professional graphical designers.

5.1. Results for repeatability

Before we can measure the repeatability of a particular detector we first had to consider typical image alterations such as image rotation and image scaling. In both cases, for each image we extracted the salient points and then we computed the average repeatability rate over the database for each detector. The repeatability rate was computed using Equation (2).

In the case of image rotation the rotation angle varied between 0° and 180° . The repeatability rate in a $\varepsilon=1$ neighborhood for the rotation sequence is displayed in Figure 2. The detectors using wavelet transform (Haar and

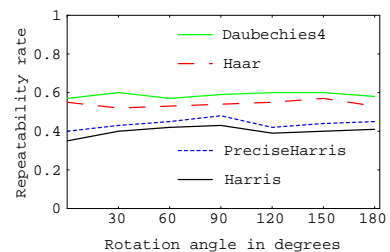


Figure 2: Repeatability rate for image rotation ($\varepsilon=1$)

Daubechies4) give better results compared with the other ones. Note that the results for all detectors are not very dependent on image rotation. The best results are provided by Daubechies4 detector.

In the case of scale changes, for each image we considered a sequence of images obtained from the original image by reducing the image size so that the image was aspect ratio preserved. The largest scale factor used was 4. The repeatability rate for scale change is presented in Figure 3.

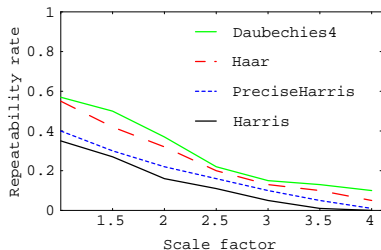


Figure 3: Repeatability rate for scale change ($\epsilon=1$)

All detectors are very sensitive to scale changes. The repeatability is low for a scale factor above 2 especially for Harris and PreciseHarris detectors. The detectors based on wavelet transform provide better results compared with the other ones.

5.2. Results for information content

In these experiments we also considered random points in our comparison. For each image in the database we computed the mean number m of salient points extracted by different detectors and then we selected m random points using a uniform distribution.

For each detector we computed the salient points for the set of images and characterized each point by a vector of local greyvalue invariants (cf. Equation (3)). The invariants were then normalized and the entropy of the distribution was computed. The cell size in the partitioning was the same in all dimensions and it was set to 20. The σ used for computing the greylevel invariants was 3.

The results are given in Table 1. This table shows that the detector using the Daubechies4 wavelet transform has the highest entropy and thus the salient points obtained are the most distinctive. The results obtained for Haar wavelet transform are almost as good. The results obtained with PreciseHarris detector are better than the ones obtained with Harris but worse than the ones obtained using the wavelet transform. Unsurprisingly, the results obtained for all of the salient points detectors are significantly better than those obtained for random points. The difference between the results of Daubechies4 and random points is about a factor of two.

Detector	Entropy
Haar	6.0653
Daubechies4	6.1956
Harris	5.4337
PreciseHarris	5.6975
Random	3.124

Table 1: The information content for different detectors

6. CONCLUSION

We presented a salient point detector based on wavelets. The wavelet-based salient points are interesting because they are located in visual focus points without gathering in textured regions. We used the Haar transform for point extraction, which is simple but may lead to bad localization. A better approach is to use Daubechies4 wavelets which avoid these drawbacks.

We also compared our detector with different corner detectors based on two criteria: repeatability and information content. In all cases the results of the wavelet-based detectors were better than those of the other detectors. All detectors have significantly more information content than randomly selected points, so they manage to select "interesting" points.

7. REFERENCES

- [1] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41:909–996, 1988.
- [2] W. Forstner. A framework for low level feature extraction. *ECCV*, 1994.
- [3] C. Harris and M. Stephens. A combined corner and edge detector. *Alvey Vis Conf*, pages 147–151, 1999.
- [4] F. Heitger, et al. Simulation of neural contour mechanism: from simple to end-stopped cells. *Vision Research*, 32(5):963–981, 1992.
- [5] R., Horaud, et al. Finding geometric and relational structures in an image. *ECCV*, 1990.
- [6] J.J. Koenderink and A.J. van Doorn. Representation of local geometry in the visual system. *Biological Cybernetics*, 55:367–375, 1987.
- [7] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Trans on Patt Anal and Mach Intell*, 19(5):530–535, 1997.
- [8] C. Schmid, et al. Evaluation of interest point detectors. *International journal of Computer Vision*, 37(2):151–172, 2000.
- [9] N. Sebe, et al. Color indexing using wavelet-based salient points. In *IEEE Workshop on Content-based Access of Image and Video Libraries*, pages 15–19, 2000.
- [10] B. ter Haar Romeny, et al. Higher order differential structure of images. *Image and Vision Computing*, 12(6):317–325, 1994.