Detecting Distinguished Regions by Saliency

Friedrich Fraundorfer and Horst Bischof

Computer Graphics and Vision, Graz University of Technology, Austria
{fraunfri,bischof}@icg.tu-graz.ac.at

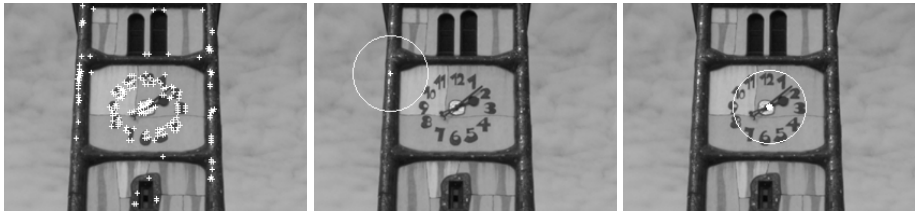
Abstract. A method for detecting and characterizing local image regions based on saliency is introduced. The proposed method detects scale localized salient regions in an image by a saliency operator which uses the concept of visual attention. A new descriptor based on a cornerness measure is presented which allows a stable identification of regions of interest and at the same time allows for an elaborate description of the identified salient regions. Experiments demonstrate that the resulting salient regions and their descriptions are discriminative enough for image matching.

1 Introduction

Recent trends in 3D-reconstruction (e.g. wide-baseline stereo) and object recognition have shown an increased use of local appearance based features and their descriptors for matching. A region of interest should contain as much information as possible. Larger regions seem to be preferable because they allow a more distinctive description but on the other hand are likely to contain occlusions if the same region is viewed from a different viewpoint. The aim therefore is to find regions which are as salient and descriptive as possible while being small enough to be completely visible from another view too. A common practice is to use a region around a corner. But if the size of the region is chosen without consideration then there is no guarantee that there are other features besides the corner, therefore the region will not be discriminative. In general, if using two different processes for identifying regions of interest (e.g. corners) and description of the region of interest (e.g. filter responses) there is no guarantee that the found regions will be discriminative. Defining a region by a high density of features inside the region is a much better choice. Fig. 1 demonstrates this idea. Fig. 1a) depicts some corners detected by the Harris operator. It is obvious from this image that the region which is best described by the Harris corners is the clock region. Fig. 1b) shows the region which is defined by the corner with the strongest response of the Harris operator and it's surrounding area with a preset diameter. The selected region is definitely not the most descriptive in this image. Fig. 1c) shows the region which is detected by the proposed method. It chooses the part of the image which is best described by Harris corners.

In the past several methods for identifying regions of interest have been developed. Some define a region by using the area surrounding a point, for instance a corner point while others define the region by finding stable borders.

J. Bigun and T. Gustavsson (Eds.): SCIA 2003, LNCS 2749, pp. 208–215, 2003.



a) b) c)

Fig. 1. a) Detected Harris corners. b) Strongest Harris corner with surrounding region. c) Best salient region.

Approaches of the first category were developed by Schmid and Mohr [1] or Lowe et al. [2]. Examples for the second approach were developed by Schaffalitzky and Zisserman [3], Mindru et.al., Tuytelaars and Van Gool [4] or Matas et.al [5].

In this paper we propose to use a method which uses the concept of visual attention. Regions of interest are detected by selecting image regions with a high density of Harris corners. The scale of the region is selected optimally by using the saliency detector of Kadir and Brady [8]. The region itself can be described in a very discriminative way by using the detected Harris corners. The Harris corners can be used for a geometric description of the region as well as a characterization of the photometric structure, e.g., using Gaussian derivatives as proposed by Schmid and Mohr [1].

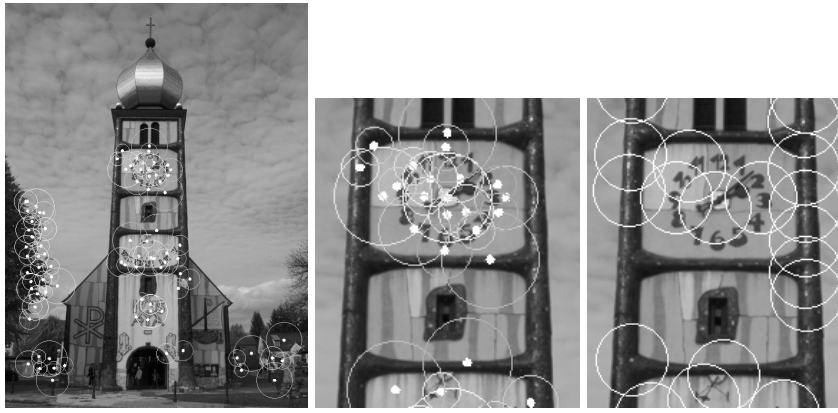
The structure of the paper is a follows. In the next section we describe the proposed region detection method. The invariance properties of the detector are discussed in section 3. In section 4 we demonstrate some properties of our new method and give an example for image matching. Finally, section 5 summarizes the paper and draws some conclusions.

2 Detecting salient regions

To overcome the decoupling of detection and description of interest regions we use a saliency operator which links this two subsequent steps. The saliency operator of Kadir and Brady allows to specify a descriptor which is used for the region detection. The basic idea of the Kadir and Brady operator is to search for clusters of high entropy in a scale-space. The operator is also able to perform automatic scale-selection. This makes it possible to choose a descriptor for the region detector which fits to the subsequent description of the regions. In this paper we describe the detected salient regions by the geometric structure of Harris corners inside of the region. Therefore we want to detect regions which contain a high number of Harris corners.

2.1 Description using cornerness values

In their work Kadir and Brady use grey-values as descriptor for the salient regions. While grey-values are the most natural kind of features they are not very



a) b) c)

Fig. 2. a) Salient regions detected by the proposed detector. b) Zoomed part of a). c) Same part as b) with regions defined by Harris corners.

suitable for description and matching of image regions. Especially correlation based matching methods are very vulnerable to illumination changes, noise or rotation and scale changes. Corner features provide a much more stable description. Our idea therefore is to use a descriptor which represents the appearance of corner features within the regions instead of the grey-values. We choose the well-known Harris corners [6] because of their high robustness [7]. The entropy calculation of the Kadir and Brady algorithm is done on the cornerness values of the image pixels. The cornerness values are calculated by $R = \det M - k(trace\ M)^2$ with

$$M = \exp^{-\frac{x^2+y^2}{2\sigma^2}} \otimes \begin{bmatrix} (\frac{\partial I}{\partial x})^2 & (\frac{\partial I}{\partial x})(\frac{\partial I}{\partial y}) \\ (\frac{\partial I}{\partial x})(\frac{\partial I}{\partial y}) & (\frac{\partial I}{\partial y})^2 \end{bmatrix} \tag{1}$$

where $I(x, y)$ is the grey level intensity and k is set to 0.04 (see the work from Harris and Stephens [6] for more details). To allow entropy calculation the cornerness values are thresholded by setting values $R < 0$ to 0 and partitioned into 256 bins from 0 to $max(R)$.

2.2 Combining different descriptors

While Kadir and Brady suggest the use of different descriptors, they do not mention the possibility to combine several different descriptors. By calculating a multi-dimensional entropy value different descriptors could be combined already in the first step. This would lead to two benefits:

- Regions which are supported by multiple descriptors will be pushed much stronger.
- Regions will be selected according to different criteria.

3 Invariancy of the saliency detector

In general detected salient regions should be invariant to a large class of transformations. Matching regions should show a high percentage of common overlapping area. Our proposed region detector is invariant to translation and rotation. This is because of the histogram calculation where the spatial information gets lost. Due to multiple window sizes scale invariance is achieved. Experiments (see section 4.2) show that while in higher resolution images naturally additional regions of interest appear (because of the more detailed image) a high number of regions are detected stable on different scales. Photogrammetric invariance depends on the used descriptor. Using the grey-values only is not invariant against illumination change but using the cornerness values of the Harris corner detector as descriptor allows an illumination insensitive description. Experiments (see section 4.3) show that the detector is also robust to viewpoint changes. The robustness depends on the used descriptor and we have seen that the Harris descriptor leads to regions which are more stable compared to using grey-values as descriptor.

4 Experimental results

4.1 Scale selection

The results of the salient region detector are compared to the results of a simple region detector with fixed size around Harris corners. Practically it is impossible to choose a suitable size for regions defined by Harris corners without using different scales. Fig. 2a) shows the results of the proposed salient region detector with a scale varying from 20 to 74 pixels in diameter. Fig. 2b) shows an enlarged part of the image. Fig. 2c) shows the same part depicting regions defined by Harris corners with a diameter of 50 pixels. It is obvious that the regions detected by our proposed method are much more similar to what a human person would identify as regions of interest. To make it worse the Harris corner method returns a lot of regions at the border of the tower which look very similar and are difficult to distinguish.

4.2 Scale invariancy

This experiment demonstrates, that matching regions can be found in images of different resolution. The scale difference was created by zooming with the camera (zoom factor 1:1.6) and not by re-sampling. 34 salient regions were detected in the lower resolution image and 46 in the higher resolution image. Thereof 19 regions were matching. Fig. 3 shows the detected regions drawn into the images and some examples of matching regions.

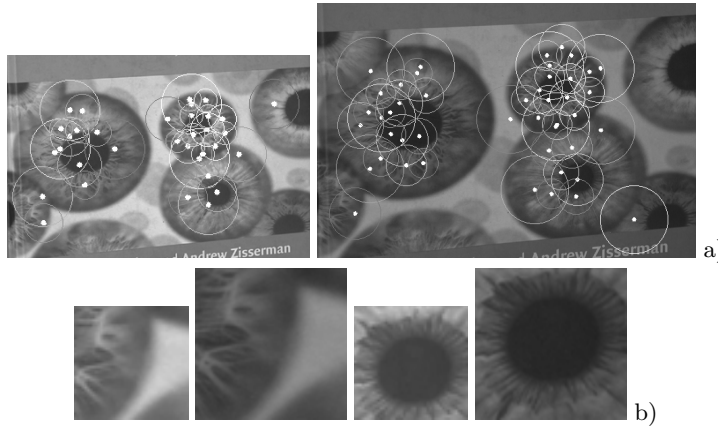


Fig. 3. a) The same scene at different scale, with the detected salient regions shown by the circles. b) Two matching regions from the scene.

4.3 Viewpoint robustness

This experiment demonstrates the robustness of the saliency operator to viewpoint changes. Fig. 4 shows the image sequence used in this experiment. The boxes were placed on a turntable and every 5 ° a picture has been taken up to a rotation of 45 °. In each image a maximum of 50 salient regions were detected using the proposed Harris descriptor. The experiment consists of detecting matching regions between the image at 0 ° and the images from the other viewpoints up to 45 °. For region matching affine invariant geometric hashing has been used as described in [9]. Table 1 shows the resulting numbers of the detected salient regions, matching regions, mis-matches and detected point correspondences. Up to the full range of viewpoint changes used in this experiment matching regions can be detected. The algorithm also detects a high number of corresponding points within the matching regions. Fig. 5 shows the detected matching regions and point correspondences for a viewpoint change of 5 °.

Viewpoint change	0 °	5 °	10 °	15 °	20 °	25 °	30 °	35 °	40 °	45 °
#detected regions	50	50	50	50	50	50	50	50	50	13
#matching regions	49	17	14	9	6	6	3	5	2	2
#mis-matches	0	0	0	0	0	0	0	0	0	0
#matching points	2686	879	721	474	308	314	154	256	105	101

Table 1. Number of matching salient regions and point correspondences for 10 different viewpoints.

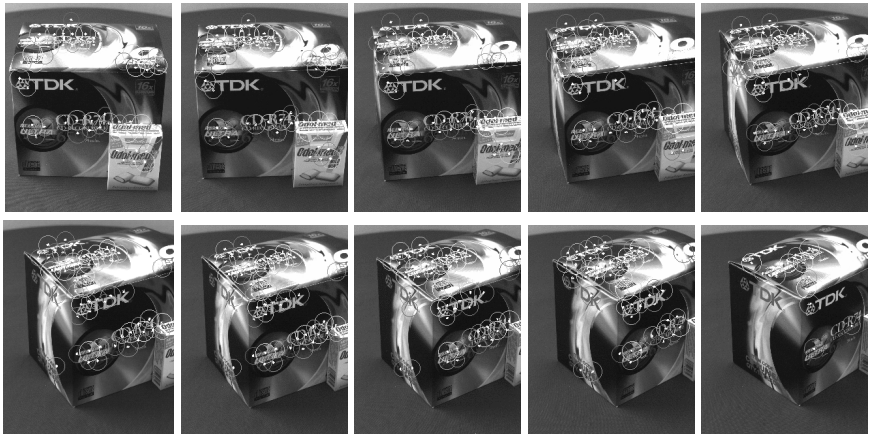


Fig. 4. Boxes scene with detected salient regions from a viewpoint of 0 ° to 45 °

Fig. 5. Detected matching regions between 0 ° to 5 °

4.4 Matching results

In this section we evaluate the usability of our salient regions for image matching and compare the proposed method to a standard approach. In every image of a stereo pair 50 salient regions were detected either by using the proposed method or by selecting regions with pre-set diameter defined by the 50 strongest Harris corners (based on the cornerness value). Fig. 6 shows the image pair and the detected salient regions using the proposed method. Region matching is done by affine invariant geometric hashing which works by comparing the geometric structure of Harris corners and the photometric structure of line features within the regions [9]. If the number of detected corresponding points within the regions is higher than a certain threshold the regions are considered to correspond. Fig. 7 shows the progress of the relative number of mis-matches based on the full set of detected regions over different thresholds. If the detected regions are discriminative and descriptive we expect only a small number of mis-matches.

Fig. 6. St. Barbara church image pair with detected salient regions using the proposed method.

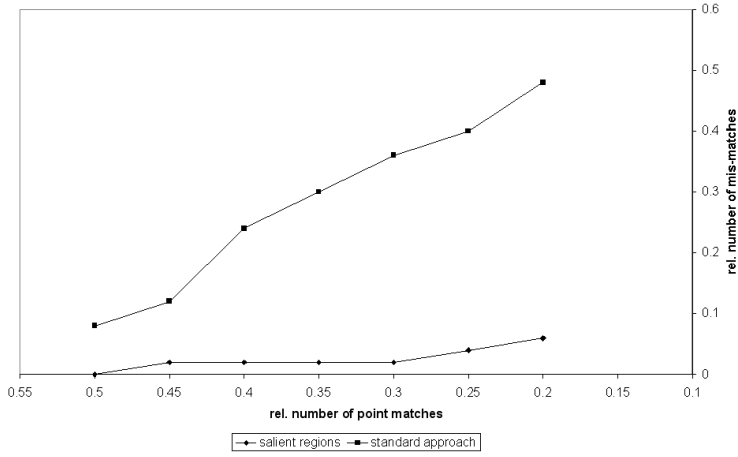


Fig. 7. Relative numbers of mis-matches over different thresholds for correspondence detection.

The graph shows that the number of mis-matches produced from our method is significantly lower compared to the standard approach. Fig. 8 shows some examples of regions detected by the standard approach with weak descriptive content. All of the examples have in common that they feature a strong corner but large parts of the region around the corner are very homogenous and therefore non descriptive. This leads to a geometric structure which is concentrated at few locations furthermore large homogenous regions without structure are non descriptive by photometric means.

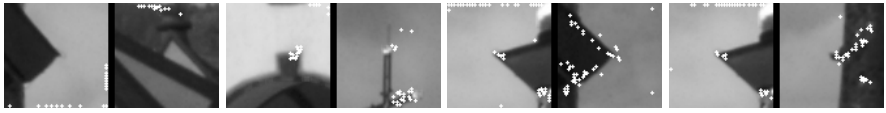


Fig. 8. Examples of regions detected (and mis-matched) by the standard approach with weak descriptive content.

5 Summary and Conclusions

In this paper we have presented a novel method for detecting and describing salient regions using the concept of visual attention. The method links the subsequent steps of detection and description of interest regions by using a common descriptor. We demonstrated this concept by introducing a descriptor based on a cornerness measure. Image matching experiments showed that salient regions detected by this concept are more descriptive and discriminative and therefore lead to better results in image matching. Especially the number of mis-matches can be reduced significantly. Furthermore by using geometric hashing for region matching (as done in this work) in addition to the information about region correspondence, point matches within these regions are established which can be used to estimate the epipolar geometry.

References

1. Schmid, C., Mohr, R.: Local grayvalue invariants for image retrieval. IEEE Trans. on Pattern Analysis and Machine Intelligence **19** (1997) 530–535
2. Lowe, D.: Object recognition from local scale invariant features. In: ICCV. (1999) 1150–1157
3. Schaffalitzky, F., Zisserman, A.: Viewpoint invariant texture matching and wide baseline stereo. In: Proc. 8th International Conference on Computer Vision, Vancouver, Canada. (2001)
4. Tuytelaars, T., Gool, L.V.: Wide baseline stereo matching based on local, affinely invariant regions. In: British Machine Vision Conference BMVC'2000. (2000)
5. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: BMVC02. (2002) 3D and Video
6. Harris, C., Stephens, M.: A combined corner and edge detector. In: Alvey Vision Conference. (1988)
7. Schmid, C., Mohr, R., Bauckhage, C.: Evaluation of interest point detectors. International Journal of Computer Vision **37** (2000) 151–172
8. Kadir, T., Brady, M.: Saliency, scale and image description. International Journal of Computer Vision **45** (2001) 83–105
9. Fraundorfer, F., Bischof, H.: Affine invariant region matching using geometric hashing of line structures. In: 27th Workshop of the Austrian Association for Pattern Recognition (OEAGM/AAPR) 2003, Laxenburg-Vienna, 2003. (2003)
10. Wolfson, H., Lamdan, Y.: Geometric hashing: A general and efficient model-based recognition scheme. In: ICCV88. (1988) 238–249
11. Manjunath, B., Ma, W.: Texture features for browsing and retrieval of image data. IEEE Trans. on Pattern Analysis and Machine Intelligence **18** (1996) 837–842