ROBUST AND FULLY AUTOMATED IMAGE REGISTRATION USING INVARIANT FEATURES

Joachim Bauer, Horst Bischof*, Andreas Klaus, Konrad Karner
VRVis Research Center, Austria
[bauer,karner]@vrvis.at
*Institute for Computer Graphics and Vision, Graz University of Technology, Austria
bischof@icg.tu-graz.ac.at

KEY WORDS: Photogrammetry, Architecture, Geometry, Matching, Feature.

ABSTRACT

This paper introduces a novel method for affine invariant matching using Zwickels that is especially well suited for images of man-made structures. Zwickels are sections defined by two intersecting line segments, dividing the neighborhood around the intersection point into two sectors. The information inside the smaller sector is used to compute an affine invariant representation. We rectify the sector using the line information and compute a histogram of the edge orientations as a description vector. The descriptor combines the advantage of accurate point localization through line intersection as well as higher descriptivity through use of a larger image region compared to descriptors computed around the points. Compared to other affine invariant descriptors we demonstrate that our method avoids the problem of depth discontinuities. In several matching experiments we show that our features are insensitive against viewpoint changes as well as illumination changes. Results are presented for aerial and terrestrial images as well.

1 INTRODUCTION

The computation of features that are invariant against viewpoint and illumination changes is a crucial step in every image matching or image indexing task. Commonly used features are the affine invariant ones, since perspective transforms, as they occur in wide baseline setups can be locally approximated by an affine transform. Typically an interest point detector provides locations at which a local affine invariant descriptor is computed. Based on the assumption, that the area around the interest point is planar or sufficiently smooth an affine invariant descriptor is useful. Several methods have been proposed in literature e.g. by. Baumberg (Baumberg, 2000), Lowe(Lowe, 1999), Schmid and Mohr (Schmid and Mohr, 1997). Mikolajczyk and Schmid (Mikolajczyk and Schmid, June 2003) evaluated the performance of several local descriptors. The most challenging problem in these approaches is to find the correct scale i.e. the spatial extension of the support region around the point. Other methods define an invariant region by finding a stable border as proposed by Schaffalitzky and Zisserman (Schaffalitzky and Zisserman, 2001), Tuytelaars and Van Gool (Tuytelaars and Gool, 2000) or Matas et.al (Matas et al., 2002). Larger regions seem to be preferable because they allow a more distinctive description, but on the other hand are more likely to contain occlusions if the same region is viewed from a different viewpoint. Larger regions may also deviate from the planar case or exhibit large perspective distortion.

In this paper we present a method for the detection and affine invariant description of image regions using Zwickels [1]. A Zwickel is formed by the intersection of two lines, where the intersection points of the line segments serve as interest points. The principal idea behind this approach is, that the area between intersecting lines is in many cases planar. Unlike other methods that compute the descriptor for a symmetric or skew-symmetric region around the interest point, we use the dividing property of the line segments to compute the descriptor only for the smaller sector. This has the advantage, that if two sectors match, we compare only the correct parts and thereby achieve a higher discrimination ability, especially if lines are lying on depth discontinuities. Our approach is split up into two steps: first we detect potential Zwickels by searching for intersecting line pairs. This step yields accurate points of interest and subdivides the region around this point into two sectors. The lines therefore automatically provide a segmentation by dividing the region around the interest point into two sectors.
In the second step we compute affine invariant descriptors for those sectors that are enclosed by the intersecting lines. The computation of the affine invariant descriptor involves a rectification of the enclosed sector and the construction of a histogram of the edge orientations. It is clear, that the proposed interest points can only be detected in images, where a sufficient number of lines is present - this is true for images containing typical man-made structures. The geometric accuracy of the intersection points is higher than those of corner based points of interest. The outline of the paper is as follows: In section 2 we describe the detection of Zwickels and the computation of the affine invariant descriptor. Section 3 shows the application of the Zwickel descriptors for image matching. Experiments with real and synthetic images are presented in section 4, concluding remarks and an outlook in section 5 close the paper.

2 ZWICKEL DETECTION AND DESCRIPTION

In the following we describe how Zwickels are detected, explain the rectification process in more detail and address the computation of the affine invariant descriptor.

2.1 Zwickel detection

The detection of Zwickels is performed as follows: In the first step 2D line segments are extracted from the image,

[1] German: *zwicken* : to nip

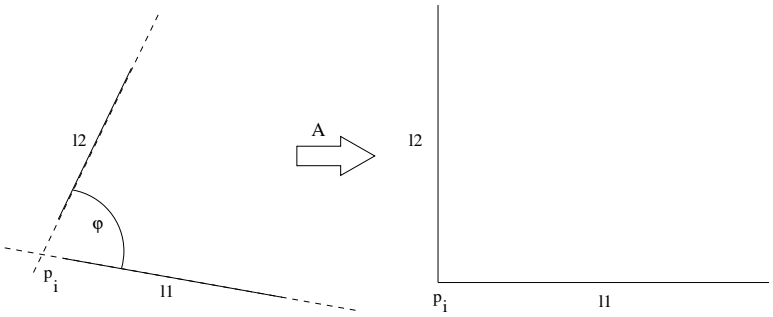
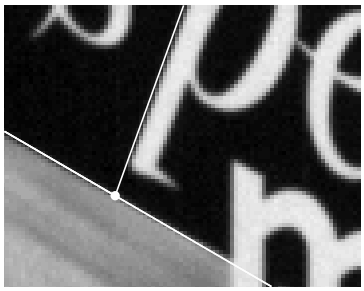
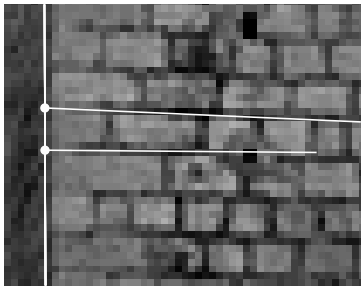


Figure 1: Left: geometry of a Zwickel: p_i is the intersection point of the lines l_1 and l_2 which are extended by a factor (extension are shown dashed) to ensure intersection. Right: For the rectification the lines l_1 and l_2 with the enclosed angle φ are mapped to an orthogonal frame using the affine transform matrix A. The transform maps the intersection point p_i to origin and the lines to the axes of the coordinate system.

those segments are extended by a predefined factor to ensure that lines, that are close enough, will intersect. All reported intersections are then handed over to the Zwickel formation procedure. For the line detection we use a hierarchical approach, that finds straight lines in a coarse-to-fine pyramid search. In every pyramid layer we extract Canny edges (Canny, 1986) with sub-pixel accuracy and fit straight line segments to sets of collinear edges. In order to compute intersections we extend the resulting line segments to ensure a sufficient number of line intersections. The detection of Zwickels is affine invariant. The lines of the detected Zwickels are ordered clockwise to ensure the correct correspondence between the lines of two matching Zwickels. As already mentioned we extend the originally extracted lines, therefore the intersection points may lie in a homogeneous region. This is one of the additional advantages over point-of-interest methods that rely on detection of location of high gradient curvature such as the Harris corner detector (Harris and Stephens, 1988). Figure 2 shows two examples of extracted Zwickels with low gradient curvature at the intersection point.



(a)



(b)

Figure 2: Examples of extracted Zwickels where the intersection point (denoted by the circle) of the two extended lines does not lie on a location of high gradient curvature i.e. no Harris corners would be detected at the intersection point.

2.2 Zwickel rectification

In order to compute an affine invariant representation of a Zwickel, we map the image data inside the sector that is bounded by the lines to an orthogonal frame (see Figure 1). An affine transform is computed from one corresponding point (the intersection point is mapped to the origin) and the two line directions. The image region in the sector is then rectified by applying the affine transform that maps the sector to an orthogonal frame with the intersection point as origin and the lines as axes of the coordinate system. Equation 1 shows the general form of an affine transform and its decomposition into a rotation, scaling and shear transform. The rectification eliminates the four of the six unknowns of the affine transform: the translation $[t_x, t_y]$ through shifting the intersection point to the origin and rotation φ and skew s through mapping the lines as orthogonal axis. The remaining unknowns are the scale factors s_x and s_y. In order to determine the unknown scale we use a similar approach as in (Lowe, 1999, Mikolajczyk and Schmid, 2001). Both approaches use a scale space search to find the correct scale of the support region.

$$
\begin{aligned}
A &= \begin{bmatrix} a_11 & a_12 \\ a_21 & a_22 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} = \qquad (1) \\
&= \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} cos\varphi & sin\varphi \\ -sin\varphi & cos\varphi \end{bmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}
\end{aligned}
$$

Figure 3 shows two examples of the rectification step.

2.3 Descriptor

In order to achieve affine invariance we apply a scale invariant descriptor. The descriptor is inspired by Lowes (Lowe, 1999) SIFT-features.

(a) (b) (c) (d)

Figure 3: Orthogonal rectification: (a) and (c) original image regions inside Zwickel. (b) and (d) rectified image regions.

We first calculate the edge orientation φ and magnitude m at each pixel inside the rectified frame I:

$$m(x,y) = \sqrt{(I_{x-1,y} + I_{x+1,y})^2 + (I_{x-1,y} + I_{x+1,y})^2} \quad (2)$$

$$\varphi(x,y) = atan((I_{x-1,y} + I_{x+1,y})/(I_{x-1,y} + I_{x+1,y})) \quad (3)$$

An orientation histogram is used as a region descriptor, the magnitude and the distance of the pixels from the origin are used as a weight. More formally the histogram is calculated as

$$H(\theta) = \sum_{\varphi \epsilon \mathcal{N}} \delta(\theta, \varphi) * w_\varphi, \quad (4)$$

where $H(\theta)$ is the value for bin θ ($\theta \in [0°, 1° \dots 360°]$) and φ denotes angle values in a neighborhood $\mathcal{N}$ inside the Zwickel, w_φ is the weight of φ and $\delta(\theta, \varphi)$ is the Kronecker delta function. The angles φ are quantized in accordance with the histogram bins θ. The weight w_φ is computed from the magnitude of φ and a function decreasing with increasing radius r from the origin (x_0, y_0). We use a Gaussian function thus $w_\varphi(x,y) = m(x,y) * g(r)$, with $r = \sqrt{(x-x_0)^2 + (y-y_0)^2}$ and $g(r) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-r^2}{2\sigma^2}}$

The parameter σ of the Gaussian function has to be adapted according to the detected scale. Due to the use of image derivatives illumination insensitivity is also achieved.

3 MATCHING

In the matching step we want to detect similar regions in an image pair. Using the Zwickel representation it is easy to implement several pre-selection criteria to speed up the matching by reducing the number of putative candidates. The pre-selection is preformed on the basis of geometric constraints as well as on image information. We only allow a maximal angle difference between corresponding lines of a Zwickel candidate pair. Furthermore we enforce the lines to have the same gradient direction. If a Zwickel encloses a darker region than the surrounding, the two lines have different gradient directions and therefore different line types.

Other pre-selection criteria for candidates e.g. by comparing the difference of the gray-value median for the Zwickels can be easily implemented. For the remaining candidates we detect the most similar ones by comparing the descriptors. In order to accomplish this task we have to choose a proper distance function for the comparison of the orientation histograms.

3.1 Distance functions

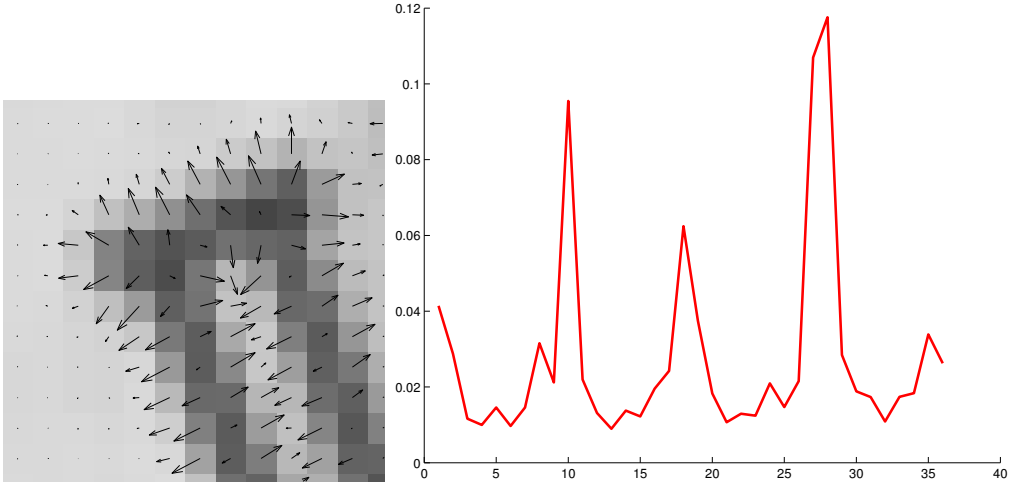
Since the descriptors described in section 4 are histograms we use probabilistic distance measures to describe the similarity. Distance measures for histogram comparison are the L_1 and L_2 norm, the Bhattacharyya distance, and the Matusita distance. The earth movers distance is a more complex method for histogram comparison and is computed by solving the so called transportation problem, proposed for image indexing by Rubner et.al (Rubner et al., 1998). Huet and Hancock (Huet and Hancock, 1996) give a comparison of the performance of this measures for histogram comparison. Following the conclusions of Rubner we chose the Bhattacharyya distance which is defined as:

$$D_{Bhatt}(H_A, H_B) = -ln \sum_i \sqrt{H_A(i) \cdot H_A(i)} \quad (5)$$

The Zwickel pair with the smallest distance is the most similar in terms of the histogram comparison.

4 EXPERIMENTS

We carried out several experiments to show the performance of the proposed method. In all experiments the region size was 30×30 pixel. In order to increase the robustness of the matching we also compute the normalized correlation coefficient cc for the rectified image patches. The distance function therefore modifies to:
$D = D_{Bhatt}(H_A, H_B) * (1 - cc(A, B))$ where A and B denote the two rectified image patches and H_A and H_B are the orientation histograms for the image patches. In the first experiment we assess the invariance of the descriptor against viewpoint changes. Sequences of several box-like objects were acquired by a turntable setup. The rotation between two subsequent images is five degrees resulting in a 72 image series. A key image is selected and we perform the matching with all subsequent images. For evaluation purposes we keep thirty percent of the best matches (smallest D)and determined the number of correct matches by calculating the epipolar geometry. Figure 6(a) and Figure 6(b) show the rate of correct matches versus the rotation angle between the camera of the key image and the camera of the second image used for the matching. The



(a) (b)

Figure 4: Visualization of orientations in the rectified frame: (a) image region with vectors visualizing the edge orientation (vector length corresponds to the magnitude). (b) histogram of edge orientations

correct matches are the inliers resulting from computing the essential matrix. The experiment is carried out with two different versions for the support region: Version one uses the sector as described in our approach. In version two the support region is centered skew symmetric around the point of interest. This comparison assesses the increase in discrimination ability when using only one sector of the interests points surrounding. The inlier rate for our approach is represented by a solid line, the dashed line is the inlier rate for the skew symmetric support region.

Figure 5 shows the differences in the used support region. Figure 6(a) shows the results for the turntable sequence

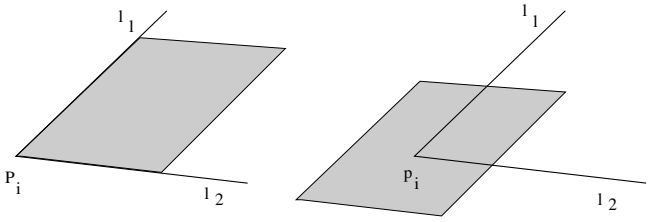
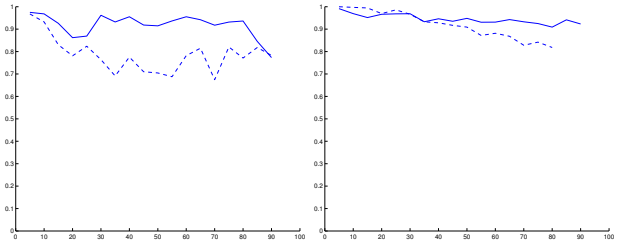


Figure 5: Illustration of the two cases for the support region. Left: support region lies inside the sector defined by the intersecting lines. Right: support region lies skew-symmetrically around intersection point

for real images. One can clearly see the superior behavior of the sector representation (approx. 20 percent increase in performance). The variance can be explained by occlusion effects e.g. when a new face of the box appears and the number of possible candidates increases or when a face disappears and the number of candidates drops. Our approach outperforms the version with the skew symmetric support region is the rotation between the cameras increases. In Figure 6(b) illustrates the results for the synthetic turntable sequence. The scene consists of a planar object with several differently structured textures 'glued' on it. Due to the lack of depth discontinuities the performance between the two versions for the support region differs less, which again nicely demonstrates the superiority of Zwickels on depth discontinuities.



(a) (b)

Figure 6: Illustration of the invariance against viewpoint changes: The rotation between the two cameras is increased in five degree steps form five to ninety degrees. The continuous line is the result for our approach, the dashed lines is for the centered support region. (a) shows the results for the data of the turntable sequence for real images. The variance results from occlusion effects, when new faces appear or other vanish. (b) illustrates the results for the synthetic turntable sequence.

In the following experiment we took several image pairs and evaluated the matching performance. Figure 7 shows the 30 percent best matching correspondences for those image pairs. Table 1 lists the results for four different image pairs. Results using other images are similar. In column 2 we list the number of total matches found, column 3 shows the number of best matching correspondences used for estimating the epipolar geometry. In column 3 and 4 we list the number of inliers and outliers accepted or rejected by enforcing epipolar consistency. Note that all image pairs show a significant rotation between views. It is clearly seen that our novel method produces many good matches and only few outliers. The matching, including the estimation of epipolar geometry, takes between 6 and

Object	total matches	matches used for epipolar geometry	inliers	outliers
aerial image pair 1	67	67	59	8
aerial image pair 2	51	51	42	9
turntable images 'Obi'	229	68	66	2
virtual turntable images	282	84	80	4
Valbonne image pair	112	50	41	9

Table 1: Evaluation of the matching performance. Results are given for 5 image pairs. Note that for the turntable images as well as for the virtual turn table scene most of the inliers lie inside a planar region, for the aerial image pairs several matches lie on depth-discontinuities where the Zwickel-based descriptor is well suited. For the Valbonne image pair several matches were found at depth discontinuities since many prominent lines were found on the borders of planar regions.

14 seconds on a Pentium 4 machine with 2.4 GHz.

5 CONCLUSION

We described a novel approach for computing affine invariant descriptors from Zwickels. Our experiments show, that these descriptors are invariant against viewpoint changes as well as illumination changes. Our method is suitable for images where a sufficient number of lines and therefore Zwickels can be extracted and the sectors inside the Zwickels provide enough texture information to distinguish competing candidates. Further possible improvements are the use of more complex distance measures for histogram comparison, such as the earth movers distance. In the next set of experiments we plan to test the method also in an object recognition context.

ACKNOWLEDGMENTS

This work has been done in the VRVis research center, Graz and Vienna/Austria (http://www.vrvis.at), which is partly funded by the Austrian government research program Kplus. Horst Bischof acknowledges the support of the Kplus competence center Advanced Computer Vision (ACV) funded by the Kplus program. The authors wish to thank Sandra Ober for providing the turntable data and Konrad Schindler for providing the virtual turntable sequence.

REFERENCES

Baumberg, A., 2000. Reliable feature matching across widely separated views.

Canny, J., 1986. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(6):679-698. 8(6), pp. 679–698.

Harris, C. and Stephens, M., 1988. A combined corner and edge detector. Proceedings 4th Alvey Visual Conference.

Huet, B. and Hancock, E., 1996. Cartographic indexing into a database of remotely sensed images. In: WACV'96, pages 8–14, Dec 1996.

Lowe, D. G., 1999. Object recognition from local scale-invariant features. In: Proc. of the International Conference on Computer Vision ICCV, Corfu, pp. 1150–1157.

Matas, J., Chum, O., Urban, M. and Pajdla, T., 2002. Robust wide baseline stereo from maximally stable extremal regions. In: Proc. 13th British Machine Vision Conference, Cardiff, UK, pp. 384–393.

Mikolajczyk, K. and Schmid, C., 2001. Indexing based on scale invariant interest points. In: Proceedings of the International Conference on Computer Vision, Vancouver, Canada, pp. 525–531.

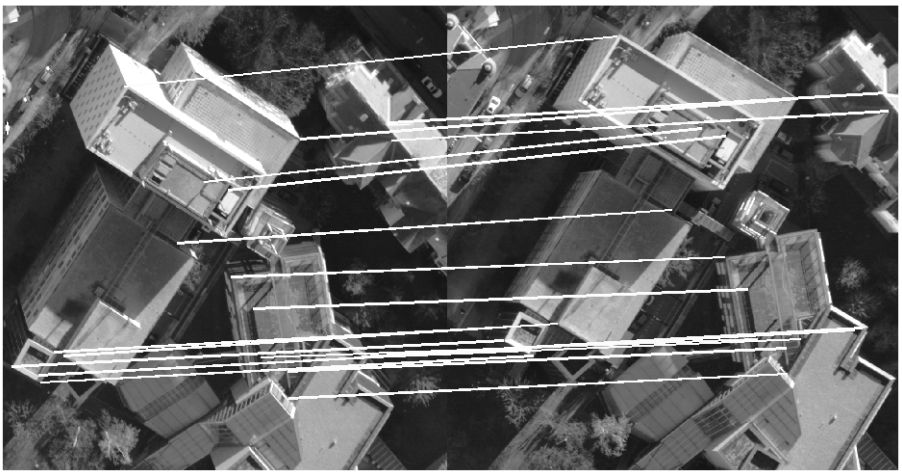
Mikolajczyk, K. and Schmid, C., June 2003. A performance evaluation of local descriptors. In: International Conference on Computer Vision and Pattern Recognition (CVPR'2003), Vol. 2, pp. 257–263.

Rubner, Y., Tomasi, C. and Guibas, L. J., 1998. A metric for distributions with applications to image databases. In: Proceedings of the 1998 IEEE International Conference on Computer Vision, Bombay, India, January 1998, pp. 59–66.

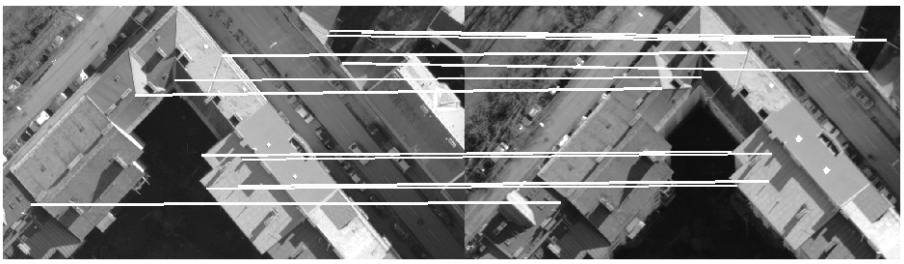
Schaffalitzky, F. and Zisserman, A., 2001. Viewpoint invariant texture matching and wide baseline stereo. In: Proc. 8th International Conference on Computer Vision, Vancouver, Canada.

Schmid, C. and Mohr, R., 1997. Local grayvalue invariants for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(5), pp. 530–535.

Tuytelaars, T. and Gool, L. V., 2000. Wide baseline stereo matching based on local, affinely invariant regions. In: British Machine Vision Conference BMVC'2000.



(a)



(b)

(c)

Figure 7: Matching results for two aerial image pairs and a terrestrial image pair (Valbonne church). For clarity only 30 percent of the best correspondences are shown. (a) Image pair1. (b) Image pair2. (c) Valbonne image pair