

Correlating Fourier descriptors of local patches for road sign recognition

Fredrik Larsson, Michael Felsberg and Per-Erik Forssen

Linköping University Post Print

N.B.: When citing this work, cite the original article.

This paper is a postprint of a paper submitted to and accepted for publication in IET Computer Vision and is subject to Institution of Engineering and Technology Copyright. The copy of record is available at IET Digital Library

Fredrik Larsson, Michael Felsberg and Per-Erik Forssen, Correlating Fourier descriptors of local patches for road sign recognition, 2011, IET Computer Vision, (5), 4, 244-254.

<http://dx.doi.org/10.1049/iet-cvi.2010.0040>

Copyright: Iet

<http://www.theiet.org/>

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-65621>

Correlating Fourier Descriptors of Local Patches for Road Sign Recognition

Fredrik Larsson Michael Felsberg Per-Erik Forssén

Computer Vision Laboratory, Department of E.E.

Linköping University, Sweden

{larsson, mfe, perfo}@isy.liu.se

July 4, 2011

Abstract

The Fourier descriptors (FDs) is a classical but still popular method for contour matching. The key idea is to apply the Fourier transform to a periodic representation of the contour, which results in a shape descriptor in the frequency domain. Fourier descriptors are most commonly used to compare object silhouettes and object contours; we instead use this well established machinery to describe local regions to be used in an object recognition framework. Many approaches to matching FDs are based on the magnitude of each FD component, thus ignoring the information contained in the phase. Keeping the phase information requires us to take into account the global rotation of the contour and shifting of the contour samples. We show that the sum-of-squared differences of FDs can be computed without explicitly de-rotating the contours. We compare our correlation based matching against affine-invariant Fourier descriptors (AFDs) and WARP matched FDs and demonstrate that our correlation based approach outperforms AFDs and WARP on real data. As a practical application we demonstrate the proposed correlation based matching on a road sign recognition task.

1 Introduction

The Fourier descriptors (FDs) [1, 2] is a classic and still popular method for contour description. The key idea is to apply the Fourier transform to a periodic representation of the contour, which results in a shape descriptor in the frequency domain. The low frequency components of the descriptor contain information about the general shape of the contour while the finer details are described in the high frequency components. Commonly, a one-dimensional parameterization of the boundary is used which enables the use of the 1D Fourier transform. Higher dimensional approaches have also been used, e.g. Generalized Fourier descriptors which describe a surface by 2-D Fourier transform [3]. Different ways for one-dimensional parameterization of the boundary, e.g. use of curvature, distance to the shape centroid, representing the boundary coordinates as complex numbers etc. have been used with FDs [4].

Traditionally FDs have been used to compare object contours. Nowadays, there are more robust (with respect to global occlusion and non-rigid deformations) but also computationally more expensive methods for comparing the global outlines of objects [5, 6, 7]. In this paper we use the well established FD machinery to describe local regions to be used in a object recognition framework, instead of describing the global outline. In general, partial occlusion of a contour makes FD-based matching fail. Using a set of local regions is more robust in this case, since all non-occluded regions still vote for the correct object. The matching step of our method is highly efficient, and thus well suited to finding tentative correspondences between two sets of detected object parts. A similar approach has been used by Lietner [8] who used modified FDs in parallel with SIFT features [9] for object recognition. We extract local regions using the Maximally Stable Extremal Regions (MSER) detector [10]. The contours of these regions are either sampled uniformly according to the affine arc length criterion, see section 2.1, or transformed with a similarity frame and then sampled in this canonical frame. We restrict the frame to similarity transformations, i.e. we roughly compensate for translation, scale and rotation, in order to keep the aspect ratio and hence to have a greater probability of separating e.g. rectangles of different aspect ratios.

In section 2 we review the theory behind Fourier descriptors. In section 3 we address the matching of FDs and explain why matching on magnitudes only is inferior to keeping the phase information. We introduce our correlation based matching scheme in section 3.2 and propose a preselection step to remove ambiguous descriptors in section 3.3. In section 4 we compare our work to the Affine-invariant Fourier descriptors (AFDs) [11] and the WARP method [12] on three datasets: Leuven, Boat, and Graf. In section 5 we demonstrate how to use FDs for road sign recognition. Finally, in section 6 we conclude and discuss future work.

2 Fourier Descriptors

In line with Granlund [1], the closed contour c with coordinates x and y is parameterized as a complex valued periodic function

$$c(l) = c(l + L) = x(l) + iy(l), \quad (1)$$

where L is the contour length, usually given by the number of contour samples.¹ By taking the 1D Fourier transform of c , the Fourier coefficients C are obtained as

$$C(n) = \frac{1}{L} \int_{l=0}^L c(l) \exp(-\frac{i2\pi nl}{L}) dl \quad n = 0, \dots, N, \quad (2)$$

where $N \leq L$ is the descriptor length.

One reason for the popularity of FDs is that they are easy to interpret. Each coefficient has a clear physical meaning and using only a few of the low frequency coefficients is equivalent to smoothing the contour. See Fig. 1 where we reconstruct a pedestrian outline starting with two low frequency coefficients and gradually add more and more high frequency components. Another strength and reason for popularity of FDs is their behavior under geometric transformations. The DC component $C(0)$ is the only one that is affected by translations c_0 of the curve $c(l) \mapsto c(l) + c_0$. By disregard-

¹We treat contours as continuous functions here, where the contour samples can be thought as of impulses with appropriate weights.

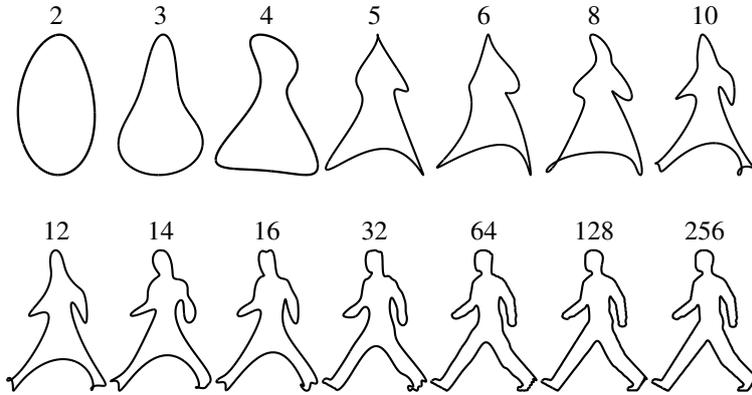


Figure 1: Reconstruction of a detail from a Swedish pedestrian crossing sign using increasing number (shown above respective contour) of Fourier coefficients.

ing this coefficient, the remaining $N - 1$ coefficients are invariant under translation. Scaling of the contour, i.e. $c(l) \mapsto ac(l)$, affects the magnitude of the coefficients and the coefficients can thus be made scale invariant by normalizing with the energy (after $C(0)$ has been removed). Without loss of generality, we assume that $\|C\|^2 = 1$ ($\|\cdot\|^2$ denotes the quadratic norm) and $C(0) = 0$ in what follows.

Rotating the contour c with ϕ radians counter clockwise corresponds to multiplication of (1) with $\exp(i\phi)$, which adds a constant offset to the phase of the Fourier coefficients

$$c(l) \mapsto \exp(i\phi)c(l) \quad \Rightarrow \quad C(n) \mapsto \exp(i\phi)C(n) . \quad (3)$$

Furthermore, if the index l of the contour is shifted by Δl , a linear offset is added to the Fourier phase, i.e. the spectrum is modulated

$$c(l) \mapsto c(l - \Delta l) \quad \Rightarrow \quad C(n) \mapsto C(n) \exp\left(-\frac{i2\pi n\Delta l}{L}\right) . \quad (4)$$

When we use the term shift we always refer to a shift in the starting point for sampling, this should not be confused with translation which we use to denote spatial translation of the entire contour.

2.1 Sampling of the Contour

We use two different approaches when sampling the contour of a region; uniform and uniform according to a first order approximation to the affine arc length [11]. In order to use the affine arc length we reparameterize the contour according to a first order approximation

$$t = \frac{1}{2} \int_l |x(l)\dot{y}(l) - y(l)\dot{x}(l)| dl . \quad (5)$$

Where $\dot{x}(l)$ and $\dot{y}(l)$ denote the derivative in the x and y directions and $x(l), y(l)$ denote the x and y coordinates. We then sample the contour at unit steps according to the new parameter t . We use a regularized derivative for estimating $\dot{x}(l)$ and $\dot{y}(l)$.

3 Matching of Fourier Descriptors

Since rotation and index-shift result in modulations of the FD, it has been suggested to neglect phase information in order to be invariant to these transformations. However, as pointed out by Oppenheim and Lim [13], most information is contained in the phase and simply neglecting it means to throw away information. Matching of magnitudes ignores a major part of the signal structure such that the matching is less specific. According to (3) and (4), the phase of each FD component is modified by a rotation of the corresponding trigonometric basis function, either by a constant offset or by a linear offset. Considering the magnitudes only can be seen as finding the optimal rotation of all different components of the FD independently. That is, given a FD of length $N - 1$, magnitude matching corresponds to finding $N - 1$ different rotations instead of estimating two degrees of freedom (constant and slope). Due to the removal of $N - 3$ degrees of freedom, two contours can be very different even though the magnitude in each FD component is the same, see Fig.2.

3.1 FD Matching Methods

Few authors made considerable efforts to really use the phase when matching FDs. In the original work [1] Granlund proposes two different methods for taking into account



Figure 2: All three contours have the same magnitude in each Fourier coefficient and the only difference is contained in the phase. A magnitude based matching scheme would return a perfect match between all of them.

the global rotation. However, there is no discussion on the effect of phase changes due to shifting the starting point, which we consider to be the more interesting problem. Persoon and Fu [14] address shifting and present a technique for estimating the least squares error for rotation, scale change and shift of the starting point. As such, their approach is closely related to ours, but they compute the minimum by numerically finding the roots of the respective derivatives of the quadratic error.

Kuhl and Giardina [15] base their matching on de-rotating the FD according to the angles estimated from the first order harmonics.² Obviously, this only works if the first harmonic locus is elliptic and in case of a circular first harmonic locus, the de-rotation requires an orientation estimate from the spatial (contour) domain: The orientation of the point with maximal distance to the center point $c(l) - c_0$ is used for de-rotation. The classification into circular and elliptic loci is obviously a matter of the noise level, i.e., the method might accidentally classify a circular domain as elliptic such that the orientation becomes arbitrary. Furthermore, very thin and lengthy structures are more or less invisible to the first order harmonics, but have a huge impact on the spatial orientation estimation variant. The pathologic case is a triangle with a very thin spike at an arbitrary position. If the triangle is equilateral, the orientation depends only on the spike, and if the triangle is slightly elongated, it is given by the largest median. Changing the triangle continuously from the former to the latter case gives a discontinuity in the orientation estimate, and thus, a poor matching result between two triangles belonging to the first and second case respectively.

²Actually this method has also been considered in [14].

Bartolini et al. have a different approach, denoted WARP, of utilizing the phase information [12]. They normalize the phase information in the descriptor (similar to [15]) and when comparing two descriptors they first use the inverse Fourier transform to reconstruct the contours. They subsequently apply *Constrained Dynamic Time Warping* (CDTW) in order to obtain a matching score for these reconstructed contours. We have adopted an implementation of CDTW by DeBarr³ in order to be able to compare our method to WARP.

Arbter et al. proposed the Affine-Invariant Fourier descriptor [11]. They keep the phase information (depending of the order) and through a product form generate a descriptor that is invariant to affine transformations. They sample the contour uniformly according to the first order approximation of the affine arc length criterion before the descriptor is extracted. This is something we have adopted and evaluated in combination with our correlation based approach. We reimplemented the work of Arbter et al. in order to be able to compare our correlation approach to the affine-invariant Fourier descriptor. We have confirmed that our AFD implementation works as intended on synthetic data. We extracted contours from one of our test images and then applied affine transformations and index shift on each contour. On these synthetic tests we got perfect precision-recall curves even under very challenging conditions such as severe foreshortening. El Oirrak et al. also propose a similar affine invariant normalization of FDs [16].

3.2 Correlation-Based FD Matching

Our approach differs in two respects from the method in [15]: First, we make use of complex FDs and avoid matrix notation. The components a, b, c, d in [15] correspond to symmetric and antisymmetric parts of the real and imaginary part of the FD. Second, we do not try to de-rotate the FDs, but we aim to find the relative rotation between two FDs, such that the matching result is maximized – similar to [14], but avoiding numerical techniques. Virtually, this is done by cyclic correlation of the contours, but due to the complex-valued FDs that we use, the same effect is achieved by multiplying

³<http://www.mathworks.com/matlabcentral/fileexchange/12319>

the FDs point-wise [17].

We show in Theorem 1 that the least-squares error between two normalized contours while compensating for rotation and index-shift can be computed by using complex correlation on FDs. In Theorem 2 we show that under the assumption that one of the contours is a transformed version of the first one, the least-squares error computed this way becomes 0. Corollary 3 shows how to obtain an estimate of the rotation, translation, index-shift and scaling that gives the least-squares error.

3.2.1 Derivation of Correlation Based Matching

Before we continue with the main result of the paper we need to formalize the normalization procedure of FDs (and corresponding contours) and also to state the complex correlation theorem.

Normalization of FDs with respect to scale and translation is achieved by

$$C(n) = \begin{cases} 0 & n = 0 \\ \frac{C'(n)}{(\sum_{n=1}^{\infty} |C'(n)|^2)^{\frac{1}{2}}} & n \neq 0 \end{cases} . \quad (6)$$

where $C(n)$ denotes the normalized FD and $C'(n)$ the general FD.

The complex correlation theorem for the periodic case is defined as [18], p. 244–245,

$$r_{12}(k) = (c_1 \star c_2)(k) \doteq \int_0^L \bar{c}_1(l) c_2(k+l) dl \quad (7)$$

$$= \sum_{n=0}^{\infty} \bar{C}_1(n) C_2(n) \exp\left(\frac{i2\pi nk}{L}\right) \quad (8)$$

$$= \mathcal{F}^{-1}\{\bar{C}_1 \cdot C_2\}(k) . \quad (9)$$

If we replace the inverse Fourier series in (8) with a truncated series of length N , we still obtain the least-squares approximation of $r_{12}(k)$.

Theorem 1. *Let \mathcal{T} denote a transformation corresponding to rotation and index-shift.*

Let c_1 and c_2 denote two normalized contours, then

$$\min_{\mathcal{T}} \|c_1 - \mathcal{T}c_2\|^2 = 2 - 2 \max_l |r_{12}(l)| \quad (10)$$

where $|\cdot|$ denotes the complex modulus and the cross correlation r_{12} is computed between the Fourier descriptors C_1 and C_2 according to (8).

Proof of Theorem 1. Taking the Fourier transform of c_1 and c_2 gives us the corresponding FDs C_1 and C_2 .

The fact that c_1 and c_2 are normalized and that \mathcal{T} corresponds to rotation and index-shift only (meaning $\|\mathcal{T}c_2\| = 1$) allows us to write the squared error between c_1 and $\mathcal{T}c_2$ as

$$\|c_1 - \mathcal{T}c_2\|^2 = (\overline{c_1 - \mathcal{T}c_2})(c_1 - \mathcal{T}c_2) \quad (11)$$

$$= \bar{c}_1 c_1 + (\overline{\mathcal{T}c_2})(\mathcal{T}c_2) - (\bar{c}_1(\mathcal{T}c_2) + (\overline{\mathcal{T}c_2})c_1) \quad (12)$$

$$= \|c_1\|^2 + \|\mathcal{T}c_2\|^2 - 2\text{Re}\{(c_1 \star \mathcal{T}c_2)(0)\} \quad (13)$$

$$= 2 - 2\text{Re}\left\{\underbrace{\exp(-i\phi)}_{\substack{\text{rotation} \\ \text{from } \mathcal{T}}}(c_1 \star c_2) \underbrace{(\Delta l)}_{\substack{\text{shift} \\ \text{from } \mathcal{T}}}\right\} \quad (14)$$

$$= 2 - 2\text{Re}\{\exp(-i\phi)r_{12}(\Delta l)\} . \quad (15)$$

It is easy to see that

$$\min_{\mathcal{T}} \|c_1 - \mathcal{T}c_2\|^2 = 2 - 2 \max_{\Delta l} |r_{12}(\Delta l)| \quad (16)$$

and the proof follows. \square

Using a finite number of Fourier coefficients means that the equality in (10) becomes an approximation.

Theorem 2. Let \mathcal{T}' and \mathcal{T} denote transformations corresponding to scaling, translation, rotation and index-shift. Let c'_1 and c'_2 denote two contours and assume that

$c'_2 = \mathcal{T}' c'_1$, then

$$\min_{\mathcal{T}'} \|c'_1 - \mathcal{T}' c'_2\|^2 = 2 - 2 \max_l |r_{12}(l)| = 0 \quad (17)$$

where $|\cdot|$ denotes the complex modulus and the cross correlation r_{12} is computed between the Fourier descriptors C_1 and C_2 according to (8).

Proof of Theorem 2. The assumption that c'_2 is a transformed version of c'_1 gives

$$c'_2(l) = s' \exp(i\phi') c_1(l - \Delta l') + t' \quad , \quad (18)$$

where ϕ' denotes the clockwise rotation in radians, $\Delta l'$ the index-shift, s' the scale change and t' the translation given by \mathcal{T}' .

Taking the Fourier transform of both contours followed by (6) results in

$$C_2(n) = \begin{cases} 0 & n = 0 \\ C_1(n) \exp(i\phi') \exp(-\frac{i2\pi n \Delta l'}{L}) & n \neq 0 \end{cases} \quad (19)$$

The cross-correlation of c_1 and c_2 via FDs is needed later, according to (9) this is given as

$$r_{12}(k) = \mathcal{F}^{-1}\{\bar{C}_1 \cdot C_2\}(k) \quad (20)$$

$$= \sum_{n=0}^{\infty} \bar{C}_1(n) \exp(i\phi') C_1(n) \exp(-\frac{i2\pi n \Delta l'}{L}) \exp(\frac{i2\pi n k}{L}) \quad (21)$$

$$= \exp(i\phi') \sum_{n=0}^{\infty} |C_1(n)|^2 \exp(\frac{i2\pi n(k - \Delta l')}{L}) \quad (22)$$

$$= \exp(i\phi') r_{11}(k - \Delta l') \quad . \quad (23)$$

Introduce a new transformation \mathcal{T}'' that corresponds only to rotation with ϕ'' and index-shift with $\Delta l''$. This gives us in the same way as before, see (11)-(15),

$$\|c_1 - \mathcal{T}'' c_2\|^2 = 2 - 2 \operatorname{Re} \{ \exp(-i\phi'') r_{12}(\Delta l'') \} \quad . \quad (24)$$

Which can be further expanded using (23), resulting in

$$\|c_1 - \mathcal{T}'' c_2\|^2 = 2 - 2\text{Re} \{ \exp(-i\phi'') \exp(i\phi') r_{11}(\Delta l'' - \Delta l') \} \quad (25)$$

The auto-correlation function r_{11} is real-valued, has its maximum which is 1 at 0 and is strictly monotonically decreasing. In order to minimize the least-squares error, select $\phi'' = \phi'$ and $\Delta l'' = \Delta l'$, which gives

$$\min_{\mathcal{T}''} \|c_1 - \mathcal{T}'' c_2\|^2 = 2 - 2 \max_l |r_{12}(l)| = 2 - 2r_{11}(0) = 0 \quad (26)$$

and the proof follows. \square

Corollary 3. *The parameters of the transformation \mathcal{T} that minimizes (17) are given as*

$$\Delta l = \arg \max_l |r_{12}(l)| \quad (27)$$

$$\phi = \arg r_{12}(\Delta l) \quad (28)$$

$$s = \frac{(\sum_{n=1}^{\infty} |C'_1(n)|^2)^{\frac{1}{2}}}{(\sum_{n=1}^{\infty} |C'_2(n)|^2)^{\frac{1}{2}}} \quad (29)$$

$$t = C'_1(0) - C'_2(0) \quad (30)$$

$$(31)$$

Proof of Corollary 3. Eq. (27) and (28) follows directly from (25) and the properties of the auto-correlation function.

Eq. (29) follows from the scale normalization that is implicitly done to both c'_1 and c'_2 by (6).

The DC-component of a FD contains the coordinates of the contour centroid. The translation from the centroid of c'_2 to the centroid of c'_1 is simply given by $C'_1(0) - C'_2(0)$. \square

3.2.2 The Correlation Based Matching Cost

Motivated by Theorem 1, the proposed correlation based matching cost between normalized contours is

$$e = 2 - 2 \max_l |r_{12}(l)| . \quad (32)$$

where the cross-correlation is computed on the corresponding FDs according to (9). Using contours normalized according to (6) and the matching cost above corresponds to approximately compensate for rotation, scaling, translation and index-shift.

Sometimes rotation invariance is not desired. The matching cost above is easily modified to only compensate for scaling, translation and index-shift by using the maximum of the real value of the complex correlation instead. That is

$$e = 2 - 2 \max_l \text{Re}\{r_{12}(l)\} \quad (33)$$

3.3 Preselection

Before we match the Fourier descriptors of regions in two images, we try to remove ambiguous descriptors. As a criterion for this we use the minimum error against all other regions in the same image e_{\min} . If $e_{\min} < T_{\text{err}}$ this particular FD is removed. The minimum error is given as

$$e_{\min} = \min_{i \neq j} \min_{\mathcal{T}} \|c_i - \mathcal{T}c_j\|^2 \quad (34)$$

where e_{\min} is estimated according to (32). Note that removing non-discriminative descriptors also reduce the computational time, as only a subset of the descriptors are kept.

3.4 Postprocessing

After having removed the ambiguous FDs within each image, we match the remaining ones between the images. Inspired by Lowe [9], we compute the error ratio e_r between

the minimum error and the second to minimum error

$$e_r = \frac{e_{\min}}{e_{\sec}} . \quad (35)$$

We use this error ratio as a way to remove insignificant matches. Experimentally we have found that a threshold of $T_r = 0.50$ returns 90% correct matches for FDs with our matching based on correlation. We do the matching in a symmetric way, i.e. we accept a match only if c_1 in image 1 matches with c_2 in image 2 and c_2 in image 2 matches with c_1 in image 1. The error ratio associated with c_1 is given as the higher one of the two error ratios.

Note that (34), (35) and the symmetric matching all reduce ambiguous matches. While being similar they are not identical and we have found that removing any of the steps will slightly degrade performance.

4 Evaluation of Matching Methods

A common approach for object recognition and pose estimation is to use local affine features. Features are extracted from views that are to be compared. These local features are commonly used in a voting scheme to find the object or pose hypothesis. We aim to use FDs in an object recognition framework, and evaluate our approach on the Leuven, Graf and Boat data set [19], see Fig. 3. These are common benchmarking sets used for testing local descriptors.

The homography relating two images in a sequence is also available. The given homographies are used solely for generating ground truth. This homography is used to estimate how one local region would be transformed into the corresponding view. We consider a reported match to be correct if it corresponds to a match given by the overlap-criterion used by Mikolajczyk et al. [19]. The subsequent precision-recall curves are generated by varying the threshold for the error ratio (35).

As mentioned earlier, we use MSER to detect local regions and two different approaches for sampling the contour. The first approach uses the affine length criterion



Figure 3: The first and third image from each of the three test sets used for evaluation.

while the second approach transforms the region into a canonical frame before sampling.

We estimate the FDs for all MSER⁴ regions in each image and compare the images pairwise. We evaluate different combinations of Fourier descriptors (Affine Fourier descriptors of order 0 and 1 (AFD0, AFD1), WARP normalized Fourier descriptors and ordinary Fourier descriptors with and without phase information FD/abs(FD)), sampling methods (Affine or Canonical) and matching methods (sum of squared differences (SSD), the proposed correlation method (Corr) Eq. 32 and constrained dynamic time warping (CDTW)). Hence, *FD Corr/Canonical* denotes ordinary Fourier descriptors with phase information sampled in the canonical frame and matched by the proposed correlation method.

The width of the Sakoe-Chiba band is set to 20 for the CDTW [12]. We use 512 sampling points on the contours for all methods and we keep the 64 Fourier coefficients corresponding to the lowest frequencies, see section 4.1 for motivation of this choice.

Figure 5 shows precision-recall curves for the three data sets. The total number of existing matches is given by the number of extracted regions that fulfills the overlap-criterion used by Mikolajczyk et al. [19]. The curves shown for the Boat and Leuven

⁴The parameters used for the MSER method were minimum margin = 30 and minimum region size = 50. We used these values for all experiments.

data sets are the cumulative results when matching the first image to the other five. We did not use the minimum error preselection criteria (34) when generating the precision-recall curves since each method would likely remove different regions. We also evaluated the performance with the preselection for a few selected combinations and the resulting precision-match curves can be seen in Fig. 6.

Apart from matching performance, the time consumption (summarized in Tables 3 and 4), is another important factor to take into consideration when choosing between different methods.

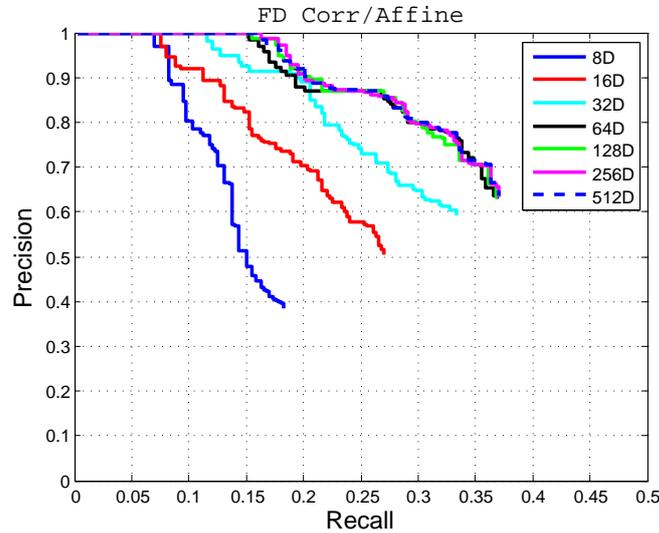


Figure 4: Precision-recall curves for varying dimensions of the FD Corr/Affine method

4.1 Different Dimensionality of FDs

The effect of varying the dimensionality of the FD by keeping different numbers of low frequency components can be seen in Fig. 4. This plot shows the precision-recall curves for the FD Corr/Affine case on the Boat data set. The performance increases with increasing dimensionality up to a certain point when reaches saturation. For the FD Corr/Affine case there seems to be no additional benefit using more than 64D. The same tendencies were shown for all evaluated methods and they all reached saturation at 64D or slightly before, which is also the reason why we have used 64D for all other

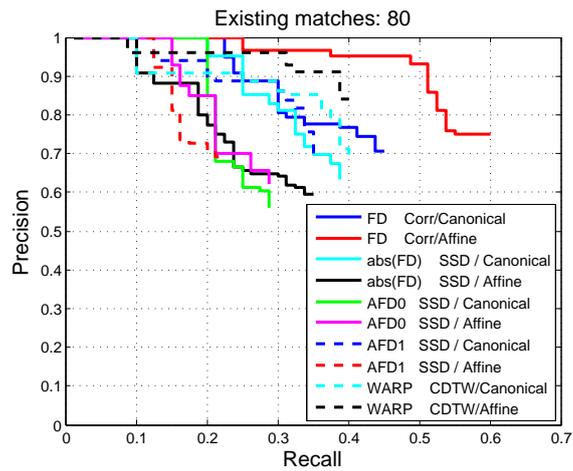
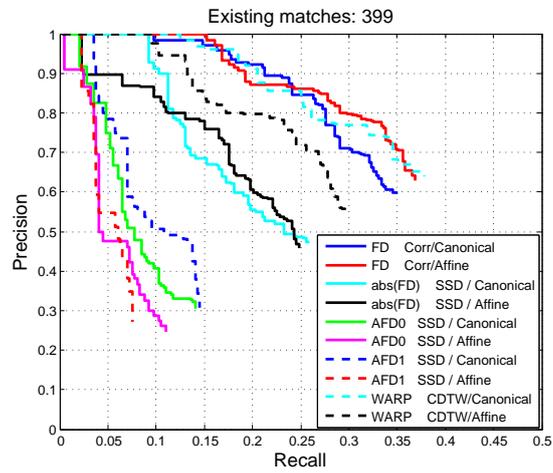
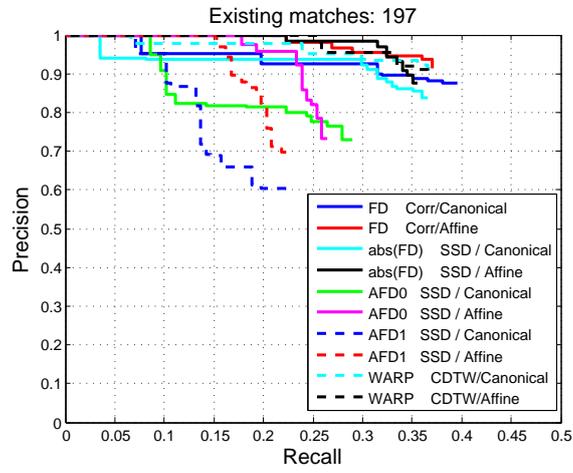


Figure 5: Precision-recall curves on the three data sets. **Top:** Leuven **Middle:** Boat **Bottom:** Graf

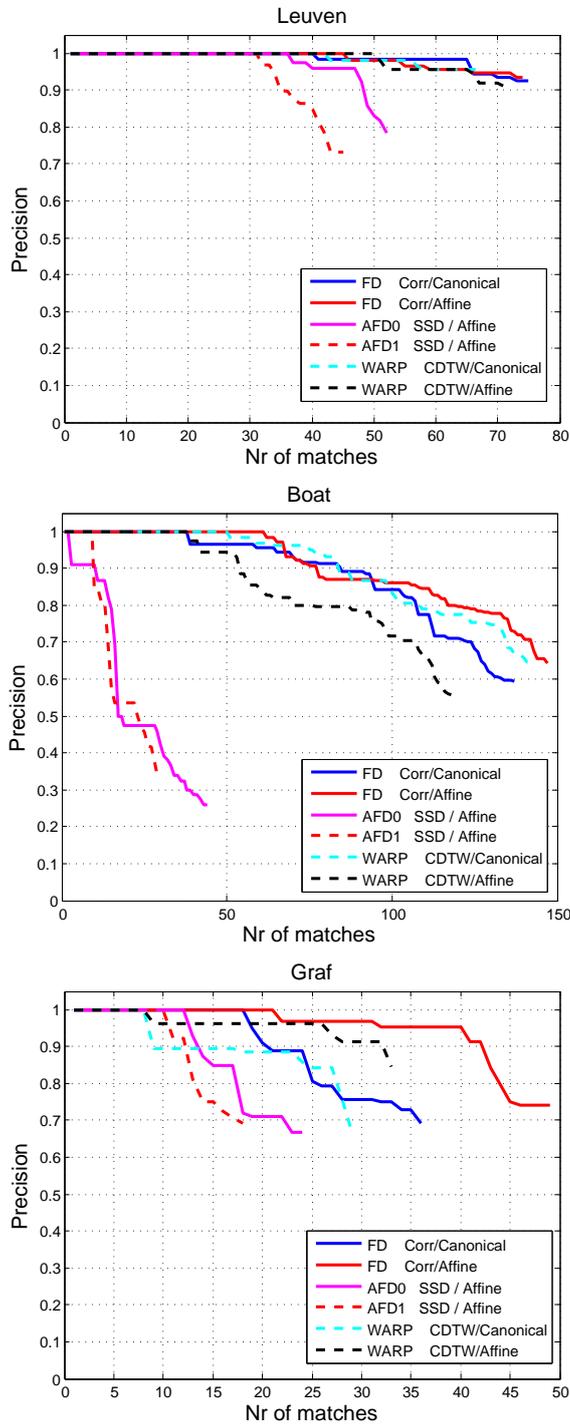


Figure 6: Precision as a function of matches. **Top:** Leuven **Middle:** Boat **Bottom:** Graf

experiments in this paper.

The number of dimensions used is not that critical, as long as one does not use too few. Using more dimensions than necessarily leads to increasing computation times with no additional benefit. The magnitude of the high frequency components tends to be relatively small, therefore adding too many does not decrease the matching performance noticeably.

Using a power-of-two number of dimensions (as well as contour sampling points) is recommended in order to benefit from the FFT speedup.

4.2 Precision-recall without preselection

Figure 5 shows the precision-recall curves for the three data sets. The area under each of the precision-recall curves, further denoted as $A(PR)$, is shown in Table 1.

	Leuven	Boat	Graf
FD Corr/Canonical	0.372	0.314	0.408
FD Corr/Affine	0.365	0.333	0.569
abs(FD) SSD/Canonical	0.341	0.200	0.354
abs(FD) SSD/Affine	0.351	0.192	0.288
AFD0 SSD/Canonical	0.251	0.086	0.259
AFD0 SSD/Affine	0.257	0.063	0.256
AFD1 SSD/Canonical	0.186	0.101	0.326
AFD1 SSD/Affine	0.212	0.054	0.203
WARP CDTW/Canonical	0.363	0.335	0.363
WARP CDTW/Affine	0.360	0.255	0.383

Table 1: Area under the precision-recall curves.

4.2.1 Precision-recall on the Leuven data set

The top plot of Fig. 5 shows the precision-recall curves for the Leuven data set. FDs in the first image of the data set are matched versus FDs from the other five images, one image at the time. Two groups emerge and the group with $A(PR) < 0.26$ contains all four versions of AFD. The differences between the methods in the top group, $A(PR) > 0.34$, are not significant. It should be noted that the Leuven dataset is supposed to test for lighting changes⁵ only. Hence, the changes in rotation between different images

⁵In reality, the camera exposure time and not scene illumination has been changed.

changes are small.

4.2.2 Precision-recall on the Boat dataset

The middle plot of Fig. 5 shows the precision-recall curve for the Boat dataset. The Boat dataset contains transformations due to zoom and rotation and we see larger differences between methods than on the Leuven dataset. We can based on precision-recall-area separate the methods into three groups. The group with $A(\text{PR}) < 0.11$ contains all AFD versions. The second best group with $0.19 < A(\text{PR}) < 0.26$ contains both magnitude only versions of the original Fourier descriptors and WARP sampled in an affine way. The best performing group, all with $A(\text{PR}) > 0.31$, contains both FD Corr methods and WARP sampled in a canonical frame. The difference in performance between FD Corr/Affine and WARP CDTW/Canonical is not significant on this test.

4.2.3 Precision-recall on the Graf dataset

The bottom plot of Fig. 5 contains the precision-recall curves for the first image pair in the Graf dataset, which corresponds to approximately 10 degrees view change.

The quality of result is divided into three levels. The lowest level with $A(\text{PR}) < 0.29$ contains abs(FD) SSD/affine and all AFD versions except from AFD1 SSD/Canonical. The top level, $A(\text{PR}) > 0.56$, consists only of the FD Corr/Affine method. For this test there is a clear distinction (difference in $A(\text{PR})$ of more than 0.15) between the top performer, FD Corr/Affine, and the second best method, FD Corr/Canonical.

For larger view changes the performance decreases but the ordering of the methods stays the same. The breakdown in performance is expected and can be explained by foreshortening effects that are not fully compensated for, despite the affine sampling.

4.3 Precision with preselection

We further evaluate the performance of incorporating the preselection criteria for AFD1 SSD/Affine, AFD0 SSD/Affine, WARP CDTW/Canonical, WARP CDTW/Affine, FD Corr/Affine and FD Corr/Canonical. We optimized the threshold for each method in-

dividually⁶. Since we remove different amounts of descriptors and also descriptors belonging to different regions for each FD method, we cannot generate fair precision-recall curves. We have instead generated precision-match curves. This allows us to see the precision but also the number of matches kept by each method. The result is also summarized in a Table 2 which shows the area under the precision-match curves, further denoted as A(PM).

	Leuven	Boat	Graf
FD Corr/Canonical	74.0	121.7	58.3
FD Corr/Affine	72.9	132.8	69.3
AFD0 SSD/Affine	50.8	25.2	41.3
AFD1 SSD/Affine	42.9	20.3	31.7
WARP CDTW/Canonical	66.3	128.8	46.8
WARP CDTW/Affine	70.8	101.7	51.9

Table 2: Area under the precision-match curves.

4.3.1 Precision with preselection on the Leuven dataset

We see in the top plot of Fig. 6 that the precision for AFD0 and AFD1 falls below 90% at around 48 and 35 matches respectively. The correlation based methods and WARP methods perform equally well and the precision of the reported matches never falls below 90%. Looking at the curves there is no real difference between the four top methods, all with $A(\text{PM}) > 66$.

4.3.2 Precision with preselection on the Boat dataset

In the middle plot of Fig. 6 we see a big difference between AFD on one side and the WARP and FD Corr methods on the other side. Both the precision and number of matches are higher for WARP and FD Corr than for AFD. We can not see any big difference between the two top methods, both with $A(\text{PM}) > 128$. For the WARP method the canonical sampling approach seems to give better results than the affine sampling approach. This relationship is the other way around for the FD Corr methods, with the affine sampling giving better result than the canonical sampling.

⁶The used thresholds are $T_{\text{err}} = 10^{-4}$ for AFD0 SSD / Affine, $T_{\text{err}} = 5 \times 10^{-4}$ for AFD1 SSD/Affine, $T_{\text{err}} = 5 \times 10^{-5}$ for WARP CDTW/Canonical, $T_{\text{err}} = 10^{-4}$ for WARP CDTW/Affine, $T_{\text{err}} = 10^{-3}$ for FD Corr/Canonical and $T_{\text{err}} = 10^{-3}$ for FD Corr/Affine.

4.3.3 Precision with preselection on the Graf dataset

The result on the Graf dataset is shown in the bottom plot in Fig. 6. The result is divided into three performance levels. The lowest level, $A(\text{PM}) < 42$ contains both AFD methods while the highest levels, $A(\text{PM}) > 69$ contains only the FD Corr/Affine method. For this dataset we see a big difference between the FD Corr/Affine method and the second best method. Interestingly, the order of the WARP curves is changed with respect to previous datasets. The affine sampling approach works better for this test for all methods.

4.4 Time Consumption

Apart from matching performance, time consumption is another important factor to take into consideration when choosing between different matching methods. Table 4 shows the time consumption for performing 10^6 matches between FDs with 64 coefficients, the time for extracting contours and creating FDs is not included. All methods are running in MATLAB except from the computations of CDTW which runs in C++. The experiment was conducted on a Intel(R) Xeon W3520 2.66 GHz CPU with 8GB RAM.

The SSD methods are the fastest ones, followed by the proposed Corr methods that are about one magnitude slower. AFD1 is complex valued, hence the added matching time compared to AFD0 and $\text{abs}(\text{FD})$. The WARP methods are the slowest ones, being two magnitudes slower than the Corr methods and three magnitudes slower than the SSD methods.

The total time for matching descriptors in two images is obviously dependent on number of regions extracted by the MSER algorithm.

The total time for matching the two first images in the Leuven data set without the preselection criteria is summarized in Table 3. For our settings 91 and 86 regions were extracted by MSER, resulting in a total of 15 652 matches since we match image 1 against image 2 and the other way around. The total time is dominated by MSER and contour sampling time for all methods except from the WARP methods for which the

actual matching time is by far the largest factor.

	MSER	Sampling ⁷	FDs	Matching	Total
FD Corr/Canonical	0.5439	0.7205	0.0260	0.2788	1.5693
FD Corr/Affine	0.5439	3.9440	0.0261	0.2799	4.7770
abs(FD) SSD/Canonical	0.5439	0.7205	0.0284	0.0220	1.3188
abs(FD) SSD/Affine	0.5439	3.9440	0.0278	0.0205	4.5362
AFD0 SSD/Canonical	0.5439	0.7205	0.1366	0.0211	1.4076
AFD0 SSD/Affine	0.5439	3.9440	0.1475	0.0222	4.6233
AFD1 SSD/Canonical	0.5439	0.7205	0.1605	0.0351	1.4636
AFD1 SSD/Affine	0.5439	3.9440	0.1617	0.0367	4.8427
WARP CDTW/Canonical	0.5439	0.7205	0.0433	21.4097	22.7106
WARP CDTW/Affine	0.5439	3.9440	0.0412	21.2876	25.7351

Table 3: Times in seconds for matching the two first images of the Leuven data set. FDs denotes time for transforming the sampled contour points into Fourier descriptors.

FD Corr/Canonical	2.938 s
FD Corr/Affine	2.991 s
abs(FD) SSD/Canonical	0.381 s
abs(FD) SSD/Affine	0.380 s
AFD0 SSD/Canonical	0.382 s
AFD0 SSD/Affine	0.382 s
AFD1 SSD/Canonical	0.821 s
AFD1 SSD/Affine	0.819 s
WARP CDTW/Canonical	670.751 s
WARP CDTW/Affine	677.544 s

Table 4: Time consumption for performing 10^6 matches with the different methods.

4.5 Summary Evaluation of Matching Methods

The top performing methods are FD Corr/Affine, FD Corr/Canonical and WARP CDTW/Canonical. The performance on the Boat and Leuven data sets are very similar for all three methods, both with and without preselection. On the Graf data set, the FD Corr/Affine method performs significantly better than the other two methods. The reason that the separation between FD Corr/Affine and FD Corr/Canonical is more distinct on the Graf data set is likely due to the fact that it is the only data is affected by foreshortening. The Leuven and Boat data sets are free from relative foreshortening.

Another important concern is the time consumption. Although achieving good results, one drawback with WARP is the added computational complexity due to CDTW.

In our implementations, with all methods running in MATLAB except from the computations of CDTW which runs in C++, the time consumption for WARP is about three magnitudes higher than for the fastest methods and two magnitudes higher than for the proposed methods. Even with further optimized implementation of CDTW the fact remains that WARP will always be slower.

The poor performance of the AFD methods on real data is obviously not caused by a faulty implementation as we have confirmed its correctness by achieving perfect results on synthetic data. We assume that this is due to the fact that in real world situations, changes in viewpoint cause changes in the extracted contours that cannot be explained by affine transformations only.

Taking all of this into account, we would recommend to use our correlation based matching method combined with affine sampling, that is FD Corr/Affine.

5 Recognizing Road Signs

As a practical application we demonstrate how to use FDs for recognizing road signs. We use synthetic images of Swedish road signs for creating models which we match against real images using the proposed correlation based matching method.

For these experiments we use the 7 signs shown in Fig. 7. The first row shows the synthetic images used to create the models and the second row shows corresponding examples from the test set. To simulate a sign detector, we extract 200 by 200 pixel patches around each sign. These patches are then processed in order to recognize any potential road signs. We use an additional 100 patches not containing any of the 7 road signs to test for false positives. About one quarter of the patches with signs contain more than a single sign.

Note that we use grey scale images and can thus not use the distinct colors occurring in the signs as a descriptor. The images used correspond to the red channel of a normal color camera. This is easily achieved by placing a red-pass filter in front of an ordinary

⁷The time difference between the different sampling schemes is likely to be less if optimizing the affine sampling. In our experiments the total running time was dominated by WARP so no time was spent on optimizing the affine sampling scheme.

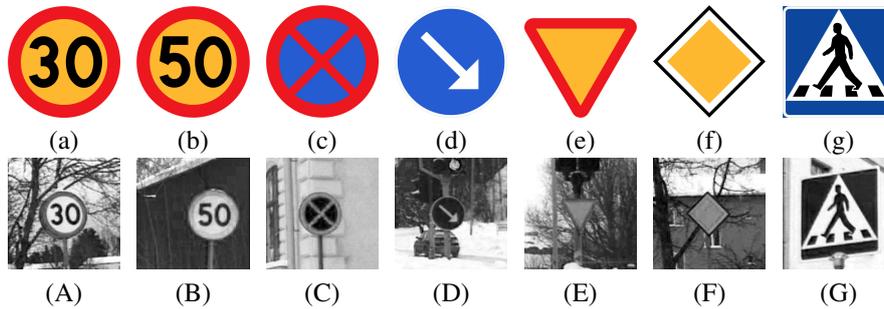


Figure 7: **First row:** Synthetic signs used to create models. **Second row:** Corresponding real world examples. **Meaning of signs:** (a,A) = 30 kph, (b,B) = 50 kph, (c,C) = No standing or parking, (d,D) = Designated lane right, (e,E) = Give way, (f,F) = Priority road, (g,G) = Pedestrian crossing.

monochromatic camera. Using normal grey-scale conversion would be problematic since some of the signs are isoluminant, e.g. sign (c) in Fig. 7. The reason for not using colors is that color cameras have lower frame rates given a fixed bandwidth and resolution. High frame rates are crucial for cameras to be used within the automotive industry. Higher frame rates mean for example higher accuracy when estimating the velocity of approaching cars.

For this particular application we do not require rotation invariance since many signs contain similar shapes but in different orientation, compare for instance the triangles in sign (e) and (g). So we use the alternate version, Eq. (33), of the proposed matching method that considers the maximum of the real value and not the absolute value.

5.1 Models

The model for each road sign consists of a subset of all contours (or rather the FDs belonging to a subset of contours) acquired by running the MSER algorithm on the synthetic image, see Fig. 8. We match all FDs in the model against the FDs extracted in the query image and accept a match if the matching cost is below an empirically set threshold. We then require that we find at least N_d out of all FDs for the model in order to say that we have found that particular sign. In general a simple form (such as a rectangle) requires a lower threshold than a more complex form (such as the pedestrian)

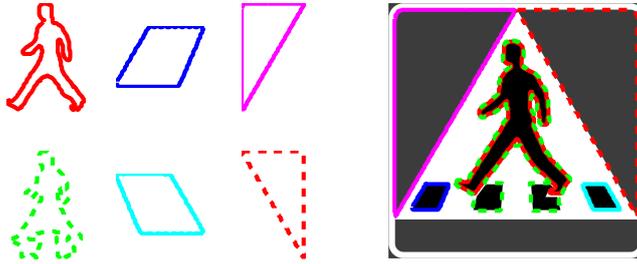


Figure 8: The contours used for the Pedestrian crossing descriptor.

Sign type:	Total	Precision	TP	FN	FP	#FDs	N_d
Pedestrian crossing	49	98.0 %	48	1	2	6	2
Designated lane right	48	95.8 %	46	2	0	1	1
No standing or parking	29	96.6 %	28	1	0	4	3
Priority road	24	91.7 %	22	2	0	1	1
Give way	24	95.8 %	23	1	0	1	1
50 kph	23	95.7 %	22	1	1	2	2
30 kph	19	89.5 %	17	2	0	2	2
All	216	95.4 %	206	10	3	-	-

Table 5: The table shows the total number of patches containing each class and the achieved number of true positives (TP), false negatives (FN) and false positives (FP). The number of FDs for each descriptor and corresponding threshold are denoted by #FDs and N_d respectively. We also used an additional 100 distractor patches to test for false positives.

to avoid too many false matches. We do not use any additional requirements such as spatial and scale proximity even though this would be an obvious extension. The reason for not including spatial information at this stage is that we want to test the performance of the pure FD matching method, which is also the main novelty of this paper.

5.2 Results

The results are summarized in Table 5. The average precision is above 95% and we have very few false positives.

The false positives acquired on the test data can be seen in Fig. 9. The matched contours are correct from a pure matching point of view, e.g. the 5 and 0 found in the left patch. Adding spatial constraints would likely remove all of these errors since the spatial relationships between the detected shapes are different from the relationships in the model.

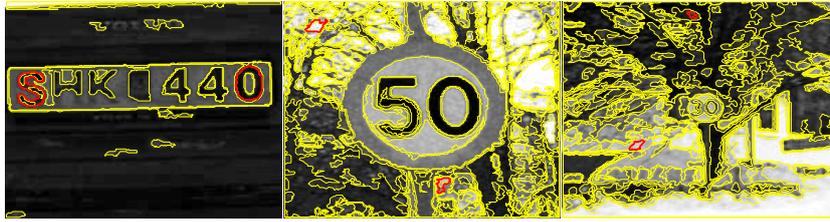


Figure 9: The false positives reported during the experiment. The yellow lines are all the extracted contours and the red lines show the contours that were assigned to the wrong prototype. **Left:** Mistaken as a 50 kph sign. **Center:** Mistaken as a pedestrian sign. **Right:** Mistaken as a pedestrian sign.

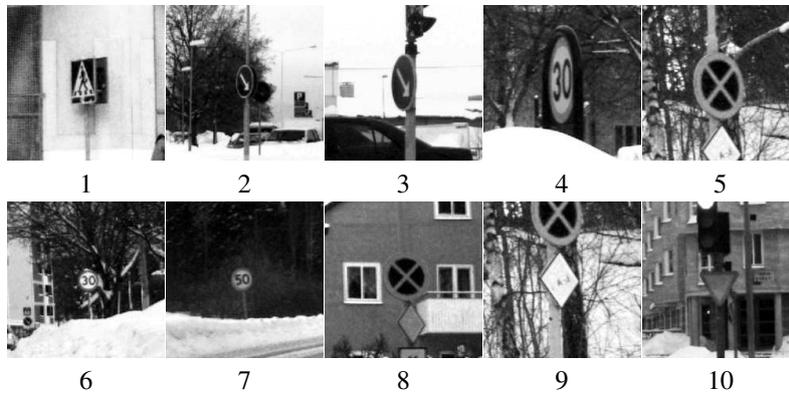


Figure 10: The missed detections during the experiment.

Fig. 10 shows the missed recognitions. The reasons for recognition failure can be divided into three types. Signs 1-5 were missed due to large out-of-plane rotations, signs 6-7 due to small signs and signs 8-10 due to failure in the contour extraction phase.

Overall, the approach of using FDs for recognizing road signs seems to be an effective approach.

6 Conclusions and Future Work

We show that the sum-of-squared differences of Fourier descriptors can be computed without explicitly de-rotating the contours using a correlation-based technique. We conclude that using Fourier descriptors to describe the shape of local regions is an efficient approach, both in matching precision-recall and in speed. Precision-recall is

significantly boosted by keeping the phase information. Computational speed benefits from the computation in Fourier domain. We suggest FDs to be used in combination with e.g. a texture descriptor, since the latter captures different aspects of the region than the FDs.

The standard approach for matching local regions is to cut out patches and describe them, e.g., using the SIFT descriptor. However, this approach has shown to be problematic when dealing with 3-D scenes with varying background [20]. For the future, we plan to apply Fourier descriptors for region matching in 3-D scenes, where the foreground patch contours are described with FDs.

We have shown that using affine sampling in combination with the proposed correlation based matching of Fourier descriptors outperforms affine invariant Fourier descriptors and WARP matched Fourier descriptors on real world data. The affine invariant Fourier descriptors achieves perfect results on synthetic data but performs poorly under real world conditions.

As a practical application we demonstrate the proposed correlation based matching on a road sign recognition task. The achieved results suggest that using FDs for road sign recognition is a promising approach. An obvious extension to the demonstrated method would be to require that the spatial and scale configuration of the detected contours are the same as for the road sign model. Also, the thresholds for each FD could be found automatically using decision theory [21].

Concluding all our experiments we see that the canonical approach is at most as good as the affine approach. We recommend to use uniform sampling according to the affine length criterion over uniform sampling in a canonical frame, because the affine sampling approach does not require the region extraction method to produce an estimate of the canonical frame. Hence, no problems with circular regions occur and a larger choice of methods for region or contour extraction is available, such as active contours.

Acknowledgment

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 215078 DIPLECS and from the Swedish Research Council through a grant for the project *Embodied Visual Object Recognition*.

References

- [1] G. H. Granlund, "Fourier Preprocessing for Hand Print Character Recognition," *IEEE Trans. on Computers*, vol. C-21, no. 2, pp. 195–201, 1972.
- [2] C. Zahn and R. Roskies, "Fourier descriptors for plane closed curves," *IEEE Transactions on Computers*, vol. C-21, no. 3, pp. 269–281, 1972.
- [3] J.-P. Gauthier, G. Bornard, and M. Silberman, "Motions and pattern analysis : Harmonic analysis on motion groups and their homogeneous spaces," *IEEE Trans. on Systems, Man and Cybernetics*, no. 21, pp. 159–178, January 1991.
- [4] A. El-ghazal, O. Basir, and S. Belkasim, "Farthest point distance: A new shape signature for Fourier descriptors," *Signal Processing: Image Communication*, April 2009.
- [5] X. Yang, S. Köknar-tezel, and L. J. Latecki, "Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2009.
- [6] P. Felzenszwalb and J. Schwartz, "Hierarchical matching of deformable shapes," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2007, pp. 1–8.
- [7] L. J. Latecki, R. Lakämper, and U. Eckhardt, "Shape descriptors for non-rigid shapes with a single closed contour," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2000, pp. 424 – 429.

- [8] R. Leitner, “Learning 3d object recognition from an unlabelled and unordered training set,” in *ISVC*, 2007, pp. 644–651.
- [9] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.
- [10] J. Matas, O. Chum, M. Urban, and T. Pajdla, “Robust wide baseline stereo from maximally stable extremal regions,” in *BMVC*, 2002, pp. 384–393.
- [11] K. Arbter, W. E. Snyder, and H. Burkhardt, “Application of affine-invariant fourier descriptors to recognition of 3-d objects,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 640–647, July 1990.
- [12] I. Bartolini, P. Ciaccia, and M. Patella, “WARP: Accurate retrieval of shapes using phase of Fourier descriptors and time warping distance,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 142–147, 2005.
- [13] A. Oppenheim and J. Lim, “The importance of phase in signals,” *Proc. of the IEEE*, vol. 69, no. 5, pp. 529–541, May 1981.
- [14] E. Persoon and K.-S. Fu, “Shape discrimination using fourier descriptors,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 7, no. 3, pp. 170–179, March 1977.
- [15] F. P. Kuhl and C. R. Giardina, “Elliptic Fourier features of a closed contour,” *Computer Graphics and Image Processing*, vol. 18, pp. 236–258, 1982.
- [16] A. El Oirrak., M. Daoudi, and D. Aboutajdine, “Affine invariant descriptors using Fourier series,” *Pattern Recognition Letters*, no. 23, pp. 1109–1118, 2002.
- [17] F. Larsson, M. Felsberg, and P.-E. Forssén, “Patch contour matching by correlating Fourier descriptors,” in *Digital Image Computing: Techniques and Applications (DICTA)*. IEEE Computer Society, 2009, pp. 40–46.
- [18] A. Papoulis, *The Fourier Integral and its Applications*. McGraw-Hill, New York, 1962.

- [19] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schafalitzky, T. Kadir, and L. Van Gool, “A comparison of affine region detectors,” *International Journal of Computer Vision*, vol. 65, pp. 128–142, 2005.
- [20] P. E. Forssén and D. Lowe, “Shape descriptors for maximally stable extremal regions,” in *IEEE International Conference on Computer Vision (ICCV)*, October 2007, pp. 1 – 8.
- [21] C. W. Therrien, *Decision, estimation, and classification: an introduction into pattern recognition and related topics*. John Wiley & Sons, Inc., 1989.