Loglet SIFT for Part Description in Deformable Part Models: Application to Face Alignment

Qiang Zhang
q.zhang.13@warwick.ac.uk
Abhir Bhalerao
abhir.bhalerao@warwick.ac.uk

Department of Computer Science
University of Warwick
Coventry, UK

Abstract

We focus on a novel loglet-SIFT descriptor for the parts representation in the Deformable Part Models (DPM). We manipulate the feature scales in the Fourier domain and decompose the image into multi-scale oriented gradient components for computing SIFT. The scale selection is controlled explicitly by tiling Log-wavelet functions (loglets) on the spectrum. Then oriented gradients are obtained by adding imaginary odd parts to the loglets, converting them into differential filters. Coherent feature scales and domain sizes are further generated by spectrum cropping. Our loglet gradient filters are shown to compare favourably against spatial differential operators, and have a straightforward and efficient implementation. We present experiments to validate the performance of the loglet-SIFT descriptor which show it to improve the DPM using a supervised descent method by a significant margin.

1 Introduction

Deformable part models (DPMs) have emerged as the leading approach for accurate landmark detection in applications such as face alignment. A DPM describes an object by local parts with a shape capturing the spatial relationships among parts. The facial landmark fitting is conducted by local feature searching followed by a shape regularisation. The performance has therefore been continually improved by employing part descriptors [16, 23] as well as shape modelling [1, 24] and fitting algorithms [15, 22, 23]. Part descriptors seek a representation of local structures which preserves intrinsic properties and discriminative information, while exhibiting invariance to changes such as illumination, scale, and variations in appearance across instances. The most successful part descriptors in DPMs are those based on oriented gradients such as SIFT [13]. The power of SIFT lies in its robustness to illumination and noise through neighbourhood pooling, and its invariance to scale achieved by salient scale selection. When SIFT descriptors are used as part "experts" in DPMs, e.g., in [22, 23, 24], the scale is selected by assigning a patch size without salience detection, therefore salient local features may not be captured. In this paper we focus on capturing wider scale ranges, so preserving richer information in SIFT descriptors. We propose multi-scale filter banks designed directly in the Fourier domain which are complementary in scale.

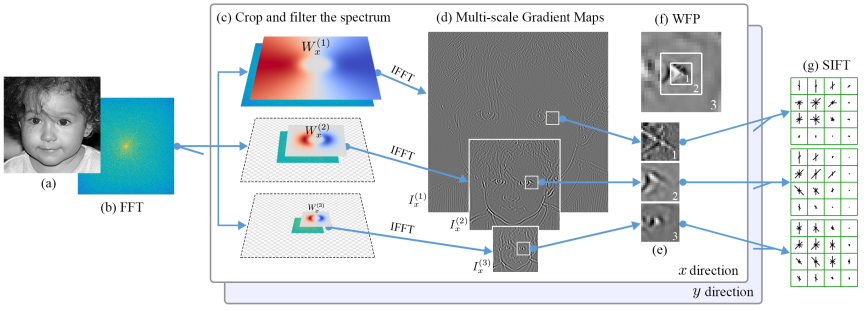


Figure 1: Overview of extracting a loglet-SIFT part descriptor.

Logarithmic wavelets (loglets) are chosen as the scale selection functions because of their superior signal processing properties [11]. Each resultant gradient map represents features at a certain scale, on which the SIFT is calculated, see an overview in Fig. 1. The new feature descriptor combines the pooling power of SIFT and scale selection of loglets and is therefore termed loglet-SIFT (L-SIFT).

Several original contributions are included in the proposed descriptor, namely: (i) We design differential filters directly in the Fourier domain with explicit scale selection; (ii) A high pass gradient filter is generated by accumulating a group of adjacent loglets, which achieves a uniform coverage towards the Nyquist frequency and is able to preserve the sharpest gradients without aliasing; (iii) Coherent feature scales and domain sizes are implemented efficiently by cropping the Fourier spectrum, which offers a more comprehensive feature descriptor, at a low computational burden.

We integrate the L-SIFT descriptor into a DPM driven by a supervised descent method (SDM) [23] and validate its performance in the face alignment scenario. We compare the performance of our Fourier domain designed filters with spatially-designed filters, and compare L-SIFT with conventional SIFT descriptors on popular face datasets. We further present the comparison against several state-of-the-art methods on two popular datasets: HELEN and 300-W. Experimental results show that L-SIFT as a part descriptor improves the performance of the DPM by a significant margin. The combined L-SIFT descriptor and SDM fitting algorithm achieves state-of-the-art performance on HELEN and 300-W common dataset, and comparable performance on the 300-W challenging dataset.

2 Related work

2.1 Multi-scale SIFT descriptors

The advantages of SIFT is its invariance to scale and illumination. However a single scale descriptor may lead to poor performance when the scale is not accurately detected [10]. In order to reduce the sensitivity to scale changes, multi-scale descriptors are proposed in feature matching scenario. For example in [19], the local feature is described with SIFT at different levels of detail within the same domain size. In [10], a set of SIFTs at multiple scales are combined for better matching performance, and in [6], a pooling across adjacent domain sizes is performed. Despite the improvement by multi-scale descriptors in feature

matching applications, the computational burden is the main obstacle when adopting them for DPMs. For example in Domain-Size Pooling (DSP) [6], scales are densely sampled and pooled (at 12 intervals in one octave) in order to marginalise the scale changes, and the computation is proportional to the number of scales. We show that pooling across adjacent scales can be approximated in the Fourier domain as filter accumulation, the implementation of which is efficient irrespective to the number of scales employed.

2.2 Wavelets

The idea of designing and tiling filters in the Fourier domain has led to efficient decomposition of local structures at multiple resolutions and orientations, e.g., steerable pyramids [20], Gabor filters [9, 14, 17], log-Gabor filters [7, 8], curvelets [21], contourlets [5], loglets [10], to name but a few. A Gabor function is a complex oscillation multiplied by a Gaussian envelope and in the Fourier domain manifests as a Gaussian function shifted away from the origin. A log-Gabor filter is a Gaussian on a logarithmic frequency scale, which has a wider bandwidth towards the higher frequencies and leads to a compact form under scaling transformations when compared with Gabor filters. A generalisation to the log-Gabor function is the loglet, as proposed by Knutsson [10], with enhanced properties such as a uniform coverage of the spectrum and an infinite number of vanishing moments (smoothness). Loglets have invariance to illumination, but because they are invariant also to sample shift they suffer less distortion caused by the limited resolution of discrete images.

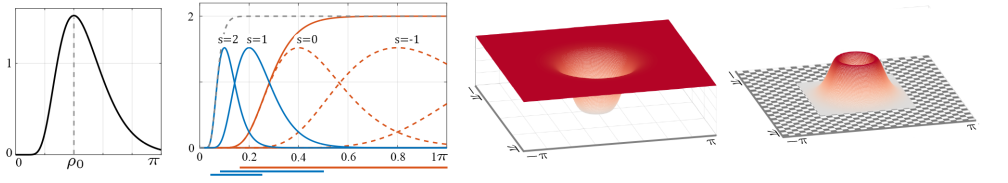
We show how loglets can be converted to differential filters to generate oriented gradients with explicit scale selection, based on the fact that all differential filters take the form of imaginary odd-windows in the Fourier space. We then design a bundle of loglets filters having a large bandwidth and covering the spectrum uniformly towards the maximum frequency, therefore the resultant gradient map preserves greater textural details than one generated by spatial filters. Moreover, we incorporate additional low pass filters for capturing information from larger scale image variation. Coherent larger domain sizes are chosen to contain these features and together they give a more comprehensive description of the local features. Our idea differs from previous wavelets-based methods in that our wavelets are designed as optimal gradient filters (imaginary odd filters) with explicit scale selection, and are further integrated into a feature descriptor such as SIFT.

3 Method

In this section we detail how to generate loglet-SIFT part descriptors for DPMs.

3.1 Feature scales

We start by decomposing an image into multiple channels with each preserving structures at certain scales. Describing the spectrum of an image in polar coordinates centred at the zero frequency, a frequency coordinate can be denoted by $\boldsymbol{u} = [\rho, \theta]$. The radius ρ actually represents a scale axis with larger scale (lower frequency) being closer to the origin. Therefore the scales can be decomposed and selected by arranging wavelets along the radius. We choose the loglets [10] as the basis functions.



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

Figure 2: Filters in the Fourier domain. (a) A loglet function. (b) A loglet filterbank. Filters at higher resolution (red dashed) are accumulated to form the first scale filter (red solid). Additional lower scale filters are shown in blue. The x coordinate, which is the radius of the polar coordinate, becomes the scale dimension. The gray dashed-line indicates the summation of all the filters, which covers the pass-spectrum uniformly. The lines at the bottom show that each filter covers octaves of the lower frequency range. (c) The 2D high pass filter. (d) The first band pass filter. The checker-board area indicates the discarded frequencies.

A loglet function is defined by,

$$\mathcal{W}(\boldsymbol{u};s)=\mathrm{erf}\left(\alpha\log\left(\beta^{s+\frac{1}{2}}\frac{\rho}{\rho_0}\right)\right)-\mathrm{erf}\left(\alpha\log\left(\beta^{s-\frac{1}{2}}\frac{\rho}{\rho_0}\right)\right) \tag{1}$$

which is a band pass filter, see Fig. 2(a). erf is an error function equals twice the integral of a normalised gaussian function. α controls the radial bandwidth, s is an integer defining the scale of the filter, and $\beta > 1$ sets the relative ratio of adjacent scales – set to two for one octave intervals. ρ_0 is the peak radial frequency of the filter with scale $s = 0$.

To preserve sharp (small scale) textures of an image, the optimal filter should cover the higher frequency components. Note that a single filter is band pass, so we need to accumulate a group of filters successively having one-octave higher central frequencies,

$$\mathcal{W}^{(1)}=\sum_{s=0,-1,\ldots}\mathcal{W}(\boldsymbol{u};s) \tag{2}$$

This achieves an even coverage towards the highest frequency benefiting from the uniformity property of loglets, see the red curve in Fig. 2(b). The resultant 2D filter is shown in Fig. 2(c). The filter accumulation enables a much larger radial bandwidth making it insensitive to scale changes. It is worth noting that the accumulation process is similar to the scale pooling used by DSP [6], where local features across adjacent spatial scales are accumulated. The reason behind the better performance of DSP is that it marginalises the feature scales, which corresponds to a wider coverage of the frequency range. This is done in our approach explicitly with much lower computation burden. We prove the equivalence of *Fourier* filter accumulation and *spatial* scale pooling under certain approximations in Appendix A in the supplementary materials.

To obtain a more comprehensive description, we extract local features at additional larger spatial scales by using filters covering the complementary lower frequency range,

$$\mathcal{W}^{(s)}(\boldsymbol{u})=\mathcal{W}(\boldsymbol{u};s-1) \tag{3}$$

Two adjacent larger scale filters at one octave intervals are shown in Fig. 2(b) as blue curves.

As the image filtering can be implemented in the Fourier domain by multiplication, the filters can be efficiently applied in the standard way,

$$I^{(s)} = \mathcal{F}^{-1}(\mathcal{I} \cdot \mathcal{W}^{(s)}), \quad s = \{1,2,...\}, \tag{4}$$

in which $\mathcal{F}$ represents the Fourier Transform and $\mathcal{I}$ the spectrum of the image I. The image is thus decomposed into multiple channels $\{I^{(s)}\}$.

3.2 Domain sizes

Given a fiducial landmark, local patches can be extracted from the image channels to obtain a multi-scale description. Larger scale textures should be described at coherently larger domain sizes and lower resolutions. We show that this is evident in the Fourier domain and can be achieved straightforwardly.

Note in Fig. 2(b) that the two larger scale filters attenuate towards high frequency and the filter magnitude beyond $\pi/2$ and $\pi/4$ is almost zero, which means little or no frequency higher than these values is preserved in the subband channels. Therefore we can cut off these areas of the spectrum, which results in an efficient image downsampling without information loss or aliasing effect.[1] With the cropping process, equation (4) becomes,

$$I^{(s)} = \mathcal{F}^{-1}(\mathcal{I}^{(s)} \cdot \mathcal{W}^{(s)}), s = \{1,2,...\}, \tag{5}$$

in which $\mathcal{I}^{(s)}$ is the cropped spectrum centred at the low frequency with $1/2^{(s-1)}$ size of the whole spectrum, $\mathcal{W}^{(s)}$ is the filter of same size as $\mathcal{I}^{(s)}$, see Fig. 1(c). As a result, the resolution of the image channels is reduced by 2^s at scale s and a subband image pyramid is obtained, see an example in Fig. 3. Note that the lowest frequency component is not covered in any of these channels as it represents the slowly varying, local mean-level containing mostly the illumination information.

At a given landmark, local patches of the same size are extracted from each of the channels, giving a multi-scale feature description (Fig. 1(e)). Although of same size in pixels, each patch represents twice the domain size and preserves one octave lower frequency components compared with its previous level. In this way a coherence between the domain size and the feature scale is achieved and the Wavelet Feature Pyramid (WFP) built (Fig. 1(f)).

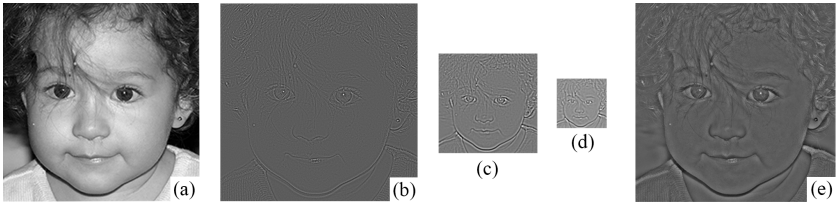


Figure 3: (a) The original image. (b)(c)(d) Pyramid of multi-scale channels with increasing scales and reducing dimensions. (e) Summation of the three channels showing the image information captured. Note that the illumination (low varying components) is suppressed as the lowest frequency band of the spectrum is discarded.

[1] Spectrum cropping as image downsampling is further explained in Appendix B in the supplementary materials.

3.3 Orientations

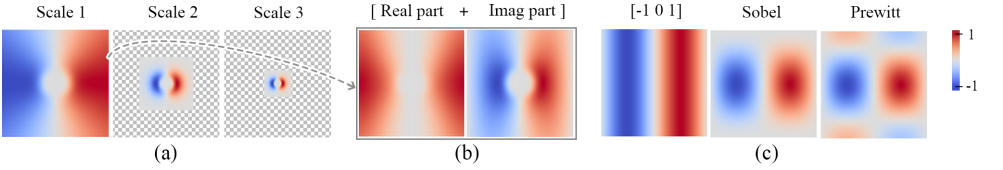


Figure 4: Filters in the Fourier domain. (a) The imaginary parts of the oriented filter banks. The real parts are zero. (b) The real and imaginary part of the first scale filter after half-pixel shift. Note that the filter is now periodically continuous. (c) For comparison, spectra (imaginary parts) of spatially defined filters.

The WFP built on multi-scale image channels can be applied to a number of intensity-based part descriptors in DPMs. Here we focus on integrating the scale selection property of loglets with the pooling power of SIFT descriptors. As SIFT performs a neighbourhood pooling on oriented gradients, we explain how to generate multi-scale gradient maps by further decomposing the non-oriented image channels into x and y components. The easiest way may seem to be by applying differential operators spatially on these channels. However the fact that differential filters take the form of an *imaginary anti-symmetrical* window in the Fourier domain (explained in appendix C in the supplementary materials), we can neatly generate the oriented gradient maps directly by converting the loglets to imaginary odd-windows.

Specifically, imaginary sinusoidal functions at orthogonal orientations are added as directional parts, decomposing the spectrum into x and y components,

$$\begin{aligned} \mathcal{W}_x^{(s)}(\boldsymbol{u}) &= j\cos(\theta)\cdot\mathcal{W}^{(s)}(\boldsymbol{u}) \\ \mathcal{W}_y^{(s)}(\boldsymbol{u}) &= j\sin(\theta)\cdot\mathcal{W}^{(s)}(\boldsymbol{u}) \end{aligned} \tag{6}$$

where θ is the orientation of vector $\boldsymbol{u}$. The oriented filters are shown in Fig. 4(a). One problem which arises is that the high pass filter (scale one) in Fig. 4(a) has larger magnitude around the Nyquist frequency (the margin of the Fourier spectrum), and its antisymmetrical shape gives $W_x(-\pi) = -W_x(\pi)$, therefore the spectrum is discontinuous across periods, which results in significant aliasing. For this reason most differential filters are designed to have zero magnitude at the boundaries to prevent aliasing, but with the penalty of losing the highest frequency components thus sacrificing precision, see Fig. 4(c). In our differential filtering, the highest frequency can be utilised without aliasing. The discontinuity is removed by adding a phase term to the odd filters,

$$\begin{aligned} \mathcal{W}_x(\boldsymbol{u}) &= e^{ju_x/2}\cdot j\cdot\cos(\theta)\cdot\mathcal{W}(\boldsymbol{u}) \\ \mathcal{W}_y(\boldsymbol{u}) &= e^{ju_y/2}\cdot j\cdot\sin(\theta)\cdot\mathcal{W}(\boldsymbol{u}) \end{aligned} \tag{7}$$

which results in a $\pi/2$ rotation in phase at one side $u_x = \pi$ and a $-\pi/2$ rotation at the other side $u_x = -\pi$, corresponding to a half-pixel shift in the spatial domain. The filters are now complex-valued and with continuity across periods, i.e., $\mathcal{W}_x(-\pi) = \mathcal{W}_x(\pi)$, see Fig. 4(b).

The gradient map $\{I_x^{(s)}, I_y^{(s)}\}$ along x and y directions at multiple scales can now be calculated by applying the oriented filters on the spectrum prior to the inverse FFT step. The

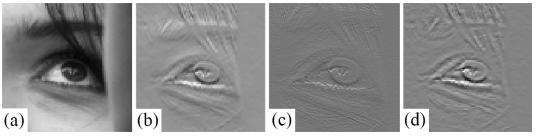


Figure 5: Illustrative comparison of differential filters. Shown are the y direction gradients produced by: (b) $[-1,0,1]$, (c) $[-1,1]$ and (d) our loglets bundle.

L-SIFT descriptor is then obtained by calculating SIFTs on the resultant multi-scale gradient maps having equal block sizes in pixels. Because larger scale channels are down-sampled, the L-SIFT features represent real domain size and scales at octave intervals.

3.4 Loglet SIFT as part experts in DPM

We integrate the L-SIFT descriptor with the SDM algorithm [23] for facial landmark detection. Denote the L-SIFT descriptors at all landmarks as $h(I,\boldsymbol{s})$, with I being the image, $\boldsymbol{s}$ the landmarks, and $h(\cdot)$ the L-SIFT extracting function. In order to deduce the true landmark location $\boldsymbol{s}^*$ given an initial estimation $\hat{\boldsymbol{s}}$, we extract the descriptor $h(I,\hat{\boldsymbol{s}})$ at $\hat{\boldsymbol{s}}$ and learn the mapping $h(I,\hat{\boldsymbol{s}}) \rightarrow \Delta\boldsymbol{s}^*$, in which $\Delta\boldsymbol{s}^* = \boldsymbol{s}^* - \hat{\boldsymbol{s}}$. The direct mapping function satisfying all the cases in the dataset is non-linear in nature and can be over-fitted. So we adopt the SDM algorithm and approximate the non-linear mapping with a sequence of linear mapping $\{R^{(i)},\boldsymbol{b}^{(i)}\}$ and landmark updating steps,

$$\begin{cases} \text{Mapping:} & \Delta\boldsymbol{s}^{(i)} = R^{(i)}h(I,\hat{\boldsymbol{s}}^{(i)}) + \boldsymbol{b}^{(i)}, \\ \text{Updating:} & \hat{\boldsymbol{s}}^{(i+1)} = \hat{\boldsymbol{s}}^{(i)} + \Delta\boldsymbol{s}^{(i)}. \end{cases} \tag{8}$$

The descriptor $h(I,\hat{\boldsymbol{s}}^{(i)})$ is extracted and updated at each iteration. Further details on SDM can be found at [23].

4 Experiments

We report the performance of the L-SIFT descriptor on the problem of face alignment with DPM. We compare our filters with spatial domain gradient filters, evaluate the improvement brought to the DPM by the proposed L-SIFT descriptors, and report the performance against state-of-the-art methods. The evaluation metric used for all the face datasets is the error normalised by the inter-pupil distance, as proposed in [2]. The parameters of the filter banks in all experiments are set as $\rho_0 = 0.3\pi$, $\alpha = 2$.

4.1 Evaluation

Comparison with other differential filters. To demonstrate the contributions of the advanced gradient filters and the multi-scale features, we first compare the single scale gradient maps generated by our first scale filter $\mathcal{W}^{(1)}$ (Fig. 4(a)) with conventional first order differential filters which can be used in SIFT descriptors, on the HELEN dataset with 68 landmarks annotated by the iBUG group. We show an example of a gradient map generated by these filters in Fig. 5. We can see that the proposed filter better preserves sharper local structures. The SIFTs are calculated on these gradients and used as the part descriptors in SDM. The results are given in Table 1. The result of the single scale filter $\mathcal{W}^{(1)}$ shows that simply replacing the conventional gradient map with the one by our filter improves the performance.

We believe this benefits from the superior properties of the loglets over spatial-designed filters, as well as the larger bandwidth achieved by the filter accumulation. We further evaluate the performance of the proposed multi-scale L-SIFT descriptor with coherent feature scales and domain sizes. The result in Table. 1 shows an additional significant improvement.

Filters	[-1 0 1]	[-1 1]	Sobel	Prewitt	$\mathcal{W}^{(1)}$	L-SIFT
Error	6.05	6.24	5.93	5.92	5.72	**5.21**

Table 1: Comparison of SIFT built on spatial filters and our filters, on Helen (68) dataset

For efficiency purposes, the filter banks can be pre-calculated and stored. The most expensive computation for generating the feature is computing the gradient maps by applying filter banks in the Fourier domain. For the single level feature, there is no additional computation comparing to a conventional SIFT based on a spatial defined operators. For a feature pyramid with s levels, the computation includes a Fourier Transform, s element-wise matrix products and inverse Fourier Transforms, both with reduced dimensions. This computation only need to be performed once before iteratively fitting the DPM to an image. Our MATLAB implementation for 3-scale features takes 9.7 ms on an image of size 400×400 using a 3.2GHz quad-core machine.

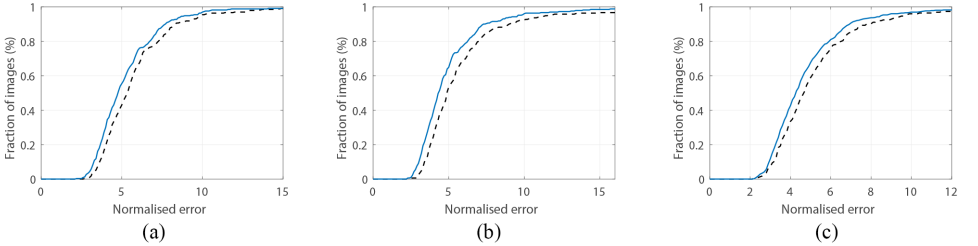


Figure 6: Improvement brought to the SDM by L-SIFT on: (a) Helen (194 landmarks), (b) Helen (68 landmarks), (c) LFPW (68 landmarks). Deshed line: SDM with SIFT; Solid line: SDM with L-SIFT.

Improvement brought to the SDM. We evaluate the improvement brought to the SDM by the L-SIFT descriptor on several datasets including the original HELEN [12] annotated with 194 landmarks, and the HELEN and LFPW dataset annotated by iBUG group using 68 landmarks. The results are shown in Fig. 6 and summarised in Table 2. We can see an improvement brought to the SDM in all datasets.

	Helen (194)	Helen(68)	LFPW(68)
SDM(SIFT)	5.85	6.05	5.32
SDM(L-SIFT)	**5.30**	**5.21**	**4.90**
Improvement	9.4%	13.9%	7.7%

Table 2: Average error of landmark fitting.

4.2 Comparison with state of the art

We compare our method with state-of-art benchmarks on the HELEN (194 points) and 300-W datasets (68 points) [18]. 300-W is created from existing datasets including LFPW, AFW,

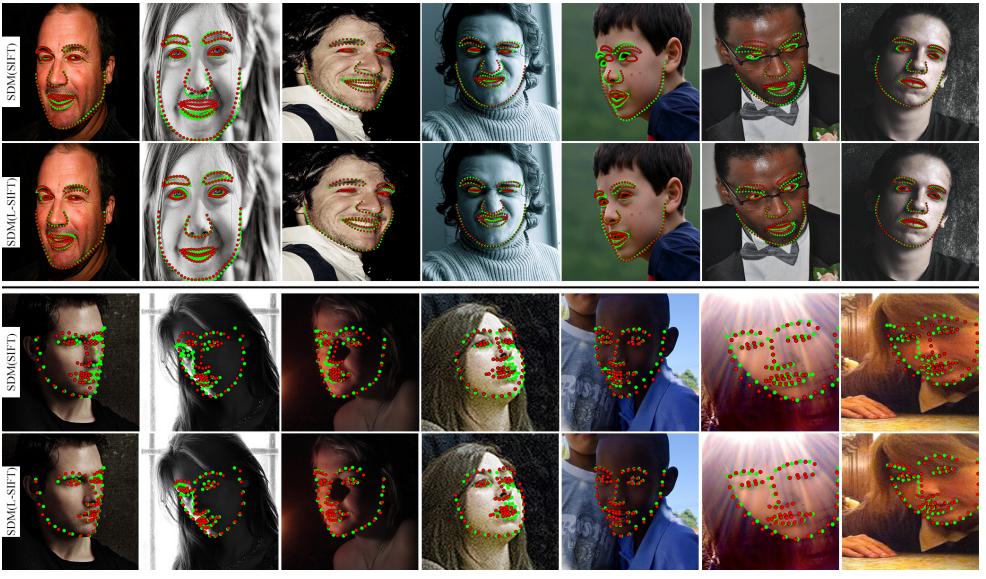


Figure 7: Qualitative results from HELEN (top row) and 300-W challenging dataset (bottom row). The SDM with L-SIFT descriptors is compared against the one with SIFT. Green points show the ground truth, and the red points the fitting results.

Method	RCPR[3]	ESR[3]	LBF fast[16]	LBF[16]	SDM(SIFT)[23]	**SDM(L-SIFT)**
Error	6.50	5.70	5.80	5.41	*5.85*	**5.30**

Table 3: Average error of methods compared on HELEN dataset

HELEN, XM2VTS and the new iBUG dataset. We follow the parameter settings given in [16]. The training set consists of AFW, the training set of LFPW and the training set of HELEN. The testing set is divided into a 'challenging' subset consisting of iBUG data and a 'common' subset consisting of the testing sets from HELEN and LFPW. The results are reported in table 3 and 4. For comparison with other methods,we list the original results in the literature.

On the HELEN dataset, the improvement by the Fourier domain designed gradient filters is more significant and the combined SDM+L-SIFT algorithm outperforms the state-of-the-art methods. On the iBUG 300-W dataset, the combined algorithm gives best results in the common subset. Although it is not as precise in the challenging subset mainly due to the large pose variations of the faces, it still improves the performance of the SDM by a useful margin. We present qualitative results on particularly challenging cases in Fig. 7 evaluating the improvement to the SDM algorithm. The results show that our feature descriptors yield better fitting performance especially on images with poor illumination or greater noise.

Method	Common Subset	Challenging Subset
ESR[4]	5.28	17.00
LBF fast[16]	5.38	15.50
LBF[16]	4.95	**11.98**
SDM(SIFT) [23]	*5.60*	*15.40*
SDM(L-SIFT)	***4.91***	*13.49*

Table 4: Average error of methods compared on 300-W dataset

5 Conclusions

This paper presents a part descriptor combining loglets and SIFT. The uniform coverage of the highest frequency gives no resolution loss and preserves the sharpest textures. Additional low frequency components are extracted, with coherently larger domain sizes achieved by cropping the Fourier spectrum, resulting in a more comprehensive feature description.

The combination of loglets and SIFT can be interpreted as an enhancement to a number of *invariances*, i.e, the invariance to illumination by the local pooling of SIFT and the suppression of slow varying mean level by the wavelets, as well as the invariances to noise by SIFT, and to sample shift by loglets. These properties improve the robustness of the descriptor to extrinsic variations. The proposed L-SIFT can be readily integrated in other gradient and SIFT based Deformable Part Models. Further work includes validating the proposed L-SIFT in computer vision tasks such as feature detection and matching. We provide a public domain version of our loglets filters and a L-SIFT toolbox for the research use, which will be made available at https://sites.google.com/site/logletsift/.

References

[1] Epameinondas Antonakos, Joan Alabort-i Medina, and Stefanos Zafeiriou. Active pictorial structures. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5435–5444, 2015.

[2] Peter N Belhumeur, David W Jacobs, David J Kriegman, and Narendra Kumar. Localizing parts of faces using a consensus of exemplars. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(12):2930–2940, 2013.

[3] Xavier P Burgos-Artizzu, Pietro Perona, and Piotr Dollár. Robust face landmark estimation under occlusion. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 1513–1520. IEEE, 2013.

[4] Xudong Cao, Yichen Wei, Fang Wen, and Jian Sun. Face alignment by explicit shape regression. *International Journal of Computer Vision*, 107(2):177–190, 2014.

[5] Minh N Do and Martin Vetterli. The contourlet transform: an efficient directional multiresolution image representation. *Image Processing, IEEE Transactions on*, 14 (12):2091–2106, 2005.

[6] Jingming Dong and Stefano Soatto. Domain-size pooling in local descriptors: DSP-SIFT. In *Computer Vision and Pattern Recognition, 2005. IEEE Conference on*, 2015.

[7] David J Field. Relations between the statistics of natural images and the response properties of cortical cells. *JOSA A*, 4(12):2379–2394, 1987.

[8] Sylvain Fischer, Filip Šroubek, Laurent Perrinet, Rafael Redondo, and Gabriel Cristóbal. Self-invertible 2D log-Gabor wavelets. *International Journal of Computer Vision*, 75(2):231–246, 2007.

[9] Markus H Gross and Rolf Koch. Visualization of multidimensional shape and texture features in laser range data using complex-valued Gabor wavelets. *Visualization and Computer Graphics, IEEE Transactions on*, 1(1):44–59, 1995.

[10] Tal Hassner, Viki Mayzels, and Lihi Zelnik-Manor. On SIFTs and their scales. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1522–1528. IEEE, 2012.

[11] Hans Knutsson and Mats Andersson. Loglets: Generalized quadrature and phase for local spatio-temporal structure estimation. In *Image Analysis*, pages 741–748. Springer, 2003.

[12] Vuong Le, Jonathan Brandt, Zhe Lin, Lubomir Bourdev, and Thomas S Huang. Interactive facial feature localization. In *Computer Vision–ECCV 2012*, pages 679–692. Springer, 2012.

[13] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[14] Oscar Nestares, Rafael Navarro, Javier Portilla, and Antonio Tabernero. Efficient spatial-domain implementation of a multiscale image representation based on Gabor functions. *Journal of Electronic Imaging*, 7(1):166–173, 1998.

[15] Chengchao Qu, Hua Gao, Eduardo Monari, Jürgen Beyerer, and Jean-Philippe Thiran. Towards robust cascaded regression for face alignment in the wild. In *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1–9. IEEE, 2015.

[16] Shaoqing Ren, Xudong Cao, Yichen Wei, and Jian Sun. Face alignment at 3000 FPS via regressing local binary features. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 1685–1692. IEEE, 2014.

[17] Yong Man Ro, Munchurl Kim, Ho Kyung Kang, BS Manjunath, and Jinwoong Kim. MPEG-7 homogeneous texture descriptor. *ETRI journal*, 23(2):41–51, 2001.

[18] Christos Sagonas, Georgios Tzimiropoulos, Stefanos Zafeiriou, and Maja Pantic. A semi-automatic methodology for facial landmark annotation. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*, pages 896–903. IEEE, 2013.

[19] Lorenzo Seidenari, Giovanni Serra, Andrew D Bagdanov, and Alberto Del Bimbo. Local pyramidal descriptors for image recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(5):1033–1040, 2014.

[20] Eero P Simoncelli, William T Freeman, Edward H Adelson, and David J Heeger. Shiftable multiscale transforms. *Information Theory, IEEE Transactions on*, 38(2): 587–607, 1992.

[21] Jean-Luc Starck, Emmanuel J Candès, and David L Donoho. The curvelet transform for image denoising. *Image Processing, IEEE Transactions on*, 11(6):670–684, 2002.

[22] Georgios Tzimiropoulos and Maja Pantic. Gauss-Newton deformable part models for face alignment in-the-wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1851–1858, 2014.

[23] Xuehan Xiong and Fernando De la Torre. Supervised descent method and its applications to face alignment. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 532–539. IEEE, 2013.

[24] Xiangxin Zhu and Deva Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2879–2886. IEEE, 2012.