Hilbert-Huang Transform-based Local Regions Descriptors

Dongfeng Han, Wenhui Li, Wu Guo
Computer Science and Technology,
Key Laboratory of Symbol Computation and
Knowledge Engineering of the Ministry of Education,
Jilin University, Changchun, P. R. China
handongfeng@gmail.com
http://handongfeng.googlepages.com/home
Zongcheng Li
School of Engineering Technology,
Shandong University of Technology, Zibo, P. R. China

Abstract

This paper presents a new interest local regions descriptors method based on Hilbert-Huang Transform. The neighborhood of the interest local region is decomposed adaptively into oscillatory components called intrinsic mode functions (IMFs). Then the Hilbert transform is applied to each component and get the phase and amplitude information. The proposed descriptors samples the phase angles information and amalgamates them into 10 overlap squares with 8-bin orientation histograms. The experiments show that the proposed descriptors are better than SIFT and other standard descriptors. Essentially, the Hilbert-Huang Transform based descriptors can belong to the class of phase-based descriptors. So it can provides a better way to overcome the illumination changes. Additionally, the Hilbert-Huang transform is a new tool for analyzing signals and the proposed descriptors is a new attempt to the Hilbert-Huang transform.

1 Introduction

Efficient local region descriptors are very useful in computer vision applications such as matching, indexing, retrieval and recognition. Commonly, the procedure of correspondences problem is not complex and includes several stages as shown in Fig.1: (1) detect the interest local regions; (2) normalize the local regions size and the main orientation; (3) compute their descriptors; (4) match the local regions using certain similar matching methods.

The object of local region detection algorithm is to specify the locations and the scales of the features. The detectors should be invariant to translation, rotation, scale, affine transform. Many scale/affine local regions detectors are proposed in the past few years [1, 2, 3, 4, 5, 6, 7]. Given invariant region detectors, the remaining question is which are the

BMVC 2007 doi:10.5244/C.21.16

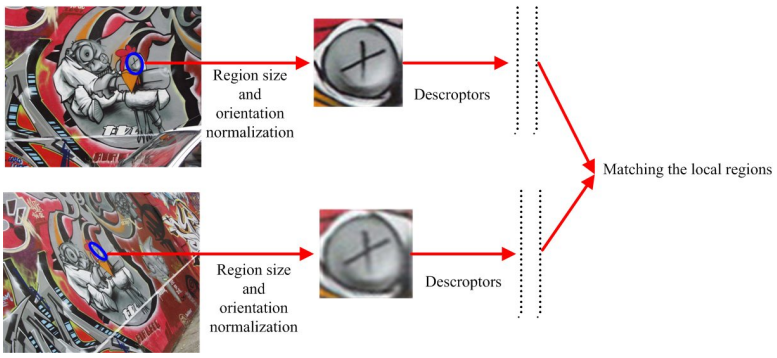


Figure 1: The main steps for the correspondences problem.

most appropriate descriptors to characterize the local regions. Many different methods for describing local interest regions have been developed [9, 10, 11, 12, 13, 14, 15, 3, 16, 8].

In this paper, a new signal process tool Hilbert-Huang transform is used to describe the local regions. The local regions are decomposed by bidimensional empirical mode decomposition (BEMD) and several intrinsic mode functions (IMFs) and the residual part are obtained. Then, Hilbert spectral analysis are conducted to describe the local regions. The Hilbert-Huang transform is locally adaptive and suitable for analysis of non-linear or non-stationary signals. Because the local regions we would process always occurs in the non-stationary regions, it would be very suitable to describe the local regions using Hilbert-Huang transform. At the same time, it shares the good features with wavelet and Fourier analyses. Also, studies in [17] shows that it can provide much better temporal and frequency resolutions than wavelet and Fourier analyses.

The experiments on standard data set show that the proposed descriptors have better results for the image illumination changes and geometry transforms. Additionally, the proposed algorithm is a new attempt to the Hilbert-Huang transform.

The rest of this paper is organized as follows. In section 2, we discuss some issues about Hilbert-Huang transform. In section 3, the Hilbert-Huang transform based local regions descriptors are described in detail. In section 4, the experiments and comparisons are given. The results of the real image experiments are demonstrated. The empirical comparisons among our approach and other five standard methods are made. The conclusions and future work are discussed in section 5.

2 Hilbert-Huang Transform

Recently, Huang et al. [17] have introduced the empirical mode decomposition (EMD) method for analyzing data. Empirical mode decomposition (EMD) is a general nonlinear, non-stationary signal processing method. So it is very suitable for describing local structure or local texture. EMD has been used in many fields such as in [18, 19, 20]. The major advantage of the EMD is that the basis functions are derived from the signal itself. Hence, the analysis is adaptive, in contrast to the wavelet method where the basis functions are fixed. The central idea of EMD is to decompose a time series into a finite and often small number of intrinsic mode functions (IMFs). As discussed in [17], an intrinsic

mode function (IMF) should satisfy two conditions: (1) in the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one; (2) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero. It is emphasized by Huang that the second condition is very important to EMD, especially for non-stationary signal (such as interest local regions). Paper [17] has a more insight on EMD.

Given a signal $x(t)$, the EMD can be represented as:

$$x(t) = \sum_{i=1}^{N} IMFs_i(t) + r(t), \tag{1}$$

where N is the number of IMFs and $r(t)$ is the residue of the signal.

Hilbert-Huang transform has the advantages in analysing the local structure. Hilbert-Huang transform includes two parts: 1) EMD; 2) Hilbert spectrum analysis (HSA). After doing EMD on the signal, some IMFs can be obtained. The next step is to analyse the $IMFs$ using HSA. The main idea of HSA is to construct complex signal for the analyzed real signal. The positive part of the spectrum of the initial real signal $x(t)$ is multiplied by two and the negative part is set to zero. Such spectrum corresponds to complex signal $z(t)$ whose imaginary part is equal to the Hilbert transform of the real part that equals to $x(t)$ as equations (2)and (3).

$$z(t) = x(t) + iy(t), \tag{2}$$

$$y(t) = H(x(t)) = v.p. \int_{-\infty}^{+\infty} \frac{x(\tau)}{\pi(t-\tau)} d\tau, \tag{3}$$

where p indicates the Cauchy principal value.

Though the above definitions are defined for any function $x(t)$ which satisfies existence conditions for the integral (3), the physical meaning of parameters phase and amplitude information is obvious only if $x(t)$ belongs to the class of monocomponent functions, i.e. the number of its extremes and the number of zero-crossings differ at most by 1 and the mean between the upper and lower envelopes equals to zero. In practice, most of the image signals are not monocomponent. By the definition of EMD in section2, the conditions are satisfied after empirical mode decomposition. EMD is very suitable for the Hilbert (or Riesz in multidimensional case) analysis.

For bidimensional signals, a similar algorithm called BEMD can be used to analysis the image signal. The principle is similar with EMD. The BEMD procedure is shown in Table1. In Fig.2, a two-level BEMD performed on a face image is shown. From Fig.2, it is clear that IMF_1 and IMF_2 have much useful information. The BEMD indeed provides the multi-scale representation of the local regions. It is shown [21] that the extracted local features have direct semantic interpretation. It contains the pattern structures from the finest to the coarsest. So the descriptors can be extracted by Hilbert analysis from IMF_1, IMF_2 and residue part.

Image is two dimensional signal and the Riesz transform [21] is a multidimensional generalization of the Hilbert transform. Because $I(x,y) = \sum_{i=1}^{n} IMF_i(x,y) + I_{r_n}(x,y)$, two dimensional complex signal can be expressed as,

$$IMF_{tA}(x,y) = IMF_t(x,y) + iIMF_{tH}(x,y), t = 1,2,...n, \tag{4}$$

Table 1: The BEMD algorithm

BEMD Algorithm:

(1) Initialization, $I_{r_0}(x,y) = I(x,y), i = 1$;

(2) Extracting ith $IMFs$:

1) Let $h_0 = I_{r_i}(x,y), k = 1$;

2) For $h_{k+1}(k)$, extracting local maxima $(m_{max,k-1})$ and minima $(m_{min,k-1})$;

3) Creating upper and lower envelope by spline interpolation of the local maxima $(m_{max,k-1})$ and minima $(m_{min,k-1})$;

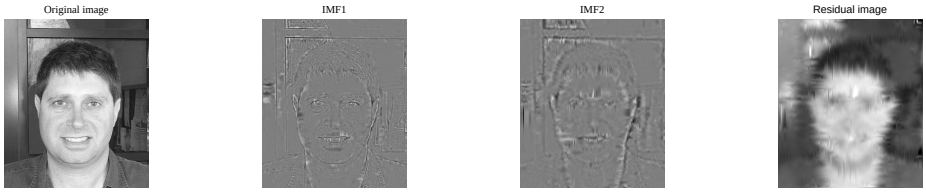
4) Computing mean value of the envelope $m_{k-1}(x,y)$;
$m_{mean,k-1}(x,y) = \frac{1}{2}(m_{max,k-1}(x,y) + m_{min,k-1}(x,y))$;

5) Computing $h_{mean,k}(x,y) = h_{k-1}(x,y) - m_{k-1}(x,y)$;

6) Checking if mean signal is close enough to zero. Yes: $h_k(x,y) = IMF_i(x,y)$, otherwise go to 2), and set $k = k+1$;

(3) $I_{r_i}(x,y) = I_{r_{i-1}}(x,y) - IMF_i(x,y)$;

(4) If the $extreme > 2$ in $I_{r_i}(x,y)$, then go to (2), and set $i = i+1$, otherwise finish. The final result:
$I(x,y) = \sum_{i=1}^{n} IMF_i(x,y) + I_{r_n}(x,y)$;



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

Figure 2: An example of BEMD. (a) The original image. (b) IMF_1 after BEMD. (d) IMF_2 after BEMD. (b) The residual image after BEMD.

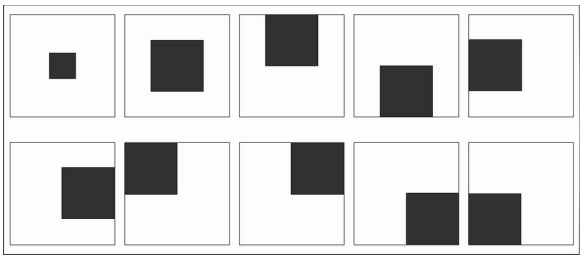


Figure 3: The illustration of overlap division for a normalized region. The local region is divided into 10 overlap regions. For each square the descriptors are computed. The formulation of division method is different from SIFT and other non-overlap methods.

where $IMF_t(x,y)$ and $IMF_{tH}(x,y)$ are called the real and complex parts of $I_{tA}(x,y)$.

So we can get the amplitude and the phase as equations (5) and (6).

$$A(x,y) = \sqrt{IMF_t^2(x,y) + IMF_{tH}^2(x,y)}, \tag{5}$$

$$\theta(x,y) = \arctan \frac{IMF_{tH}(x,y)}{IMF_t(x,y)}. \tag{6}$$

It can be denoted in a compact expression,

$$G(x,y) = A(x,y).exp(i\theta(x,y)). \tag{7}$$

In order not to lose information, the residual part as equation(8) can be analyzed by the same way.

$$I_{r_nA}(x,y) = I_{r_n}(x,y) + iI_{r_{nH}}(x,y). \tag{8}$$

Commonly, nature images always have complex texture structure and the local regions we are interested always occur in these regions. Similar with $1D$ signal, image signal is suitably analyzed by Hilbert-Huang transform.

3 Hilbert-Huang Transform Based Local Region Descriptors

Though the phase contains more information than amplitude, the phase is sensitive to the image transforms. In order to overcome this disadvantage, we project the phase onto eight orientations for each pixel as in SIFT [3].

DOG detector is used to detect local regions over scales. The characteristic scale determines the structure of the local region. The local region is rotated to its main orientation and then is normalized to 41×41. The local region is performed by two levels BEMD which can describe the detail local structure enough. Together with the residual image, the proposed algorithm will get 3 "images"(IMF_1, IMF_2, and I_{rn}).

In order to reduce the feature dimensions and maintain the spatial information , the local region is spatially divided into 10 overlap square regions. This division method can

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

Figure 4: Some test images. (a) graf image.(b) boat image. (c) car image. (d) motorcycle image.

get better result than the non-overlap in [3]. The division scheme can provide high overlap ratio with enough descriptors dimensions. In Fig.3, a division example is illustrated. In each square region, the phase angle $\theta(x,y)$)is projected onto eight orientations with the amplitude $A(x,y)$ at each position. In order to restrain the sensitivity to illumination changes, the amplitude is controlled using equation(9) as in [22]. By this way, the saturated amplitude is roughly constant for large amplitude. With the eight orientations, 10 squares and 3 images(IMF_1, IMF_2, and I_{rn}), the total dimensions of the descriptors are 240.

$$\widetilde{A}(x,y) = 1 - \exp\frac{-A^2(x,y)}{2}. \tag{9}$$

In practice, it is not necessary to conduct BEMD for every local region. In fact, the BEMD can be performed on the whole image at the beginning of the algorithm. This tip can speed up the running time.

4 Experiments and Comparisons

4.1 Matching Strategy and Comparison Criterion

Matching method is important to the final performance. For different problems, different matching methods should be used. Nearest neighbour matching and ratio matching are often used.
(1) Nearest neighbour matching: A and B are matched if the descriptor D_B is the nearest neighbor to D_A and if the distance between them is below a threshold. With this approach a descriptor has only one match.
(2) Ratio matching: this method is similar to nearest neighbor matching except that the thresholding is applied to the distance ratio between the first and the second nearest neighbour. Thus the regions are matched if $\|D_A - D_B\| / \|D_A - D_C\| < t$ where D_B is the first and D_C is the second nearest neighbour to D_A.

In this paper the ration matching method is used. We use *recall* and *precision* as the performance evaluation metric. They are defined as, $recall = \frac{\#correctmatches}{\#correspondences}$ and $precision = \frac{\#correctmatches}{\#correctmatches+\#false}$.

Two points v_i and v_j are a pair of correct match if the error in relative location is less than 3 pixels, which means v_i and v_j should satisfy $L_{v_i} - HL_{v_j} < 3$, Where H is the homography between v_i and v_j.

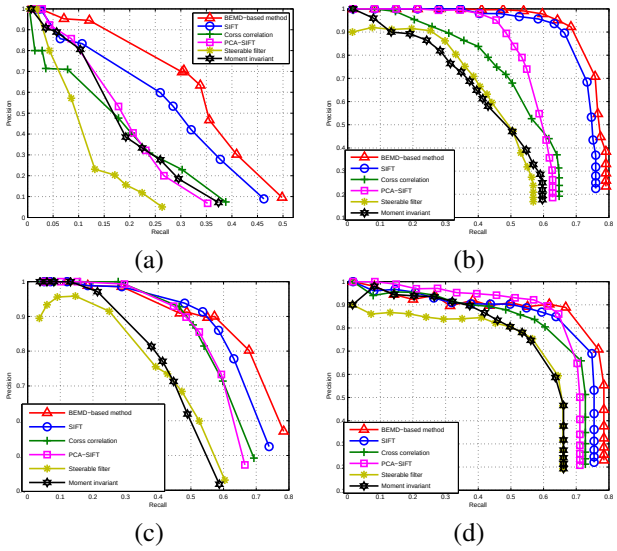


Figure 5: Performances evaluation for several image changes. (a)ROC curve under affine changes for graf image. (b) ROC curve under scale + rotation changes for boat image. (c) ROC curve under illumination changes for car image. (d) ROC curve under blur changes for motorcycle image.

4.2 Experiments Comparisons

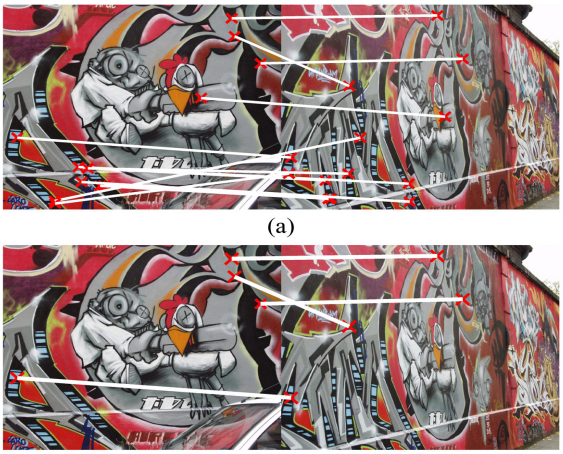
The data set is available from VGG Lab [23]. Some of the images are shown in Fig.4. We compare six different descriptors which include: 1) the proposed algorithm; 2) SIFT [3]; 3) cross correlation; 4) PCA-SIFT [16]; 5) steerable filters [11]; 6) moment invariant [9]. We generate the *recall* vs *precision* graph by changing the threshold for six different descriptors.

In Fig.5, the ROC curves are demonstrated for different image changes which include affine changes(graf image), scale+rotation(boat image), illumination changes(car image) and blur changes(motorcycle image). In our experiments, the proposed descriptors show its superior for illumination changes and affine transform. The reason is that the proposed descriptors is a phased-based descriptors essentially. For the illumination changes, the phase information (corresponding to the orientation) is more robust than other methods.

In Fig.6, a matching example is given for view point change. We compare the proposed descriptors with SIFT. It can be seen from the example that the proposed algorithm can match more accurately than SIFT.

5 Conclusions

In this paper, a new local region descriptors is presented. The main contribute of this paper is introducing Hilbert-Huang transform to local descriptors. We elaborately design an efficient compact local descriptors based on the BEMD and Hilbert transform. The main idea of the proposed descriptors is to analyze the IMFs by Hilbert spectrum. The corre-



(a)

(b)

Figure 6: Matching comparison proposed method with SIFT. The mismatches are drawn in white lines with red cross. (a) The mismatches obtained by standard SIFT. SIFT: 13/50 mismatches. (b) The mismatches obtained by the proposed algorithm. Proposed Method: 4/50 mismatches.

sponding experiments are promising. The new descriptors show its advantages especially for illumination changes and common geometry transforms.

References

[1] T. Lindeberg.: Feature detection with automatic scale selection. International Journal of Computer Vision, Vol. 30, 1998), No. 2, pp. 79–116.

[2] C. S. K. Mikolajczyk.: Indexing based on scale invariant interest points. 8th International Conference on Computer Vision, Institute of Electrical and Electronics Engineers Inc, 2001, pp. 525–531.

[3] D. G. Lowe.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, Vol. 60, 2004, No. 2, pp. 91–110.

[4] T. Tuytelaars. and L. V. Gool.: Wide baseline stereo matching based on local, affinely invariant regions. The Eleventh British Machine Vision Conference, 2000, pp. 412–425.

[5] T. Tuytelaars and L. Van Gool.: Matching widely separated views based on affine invariant regions. International Journal of Computer Vision, Vol. 59, 2004, No. 1, pp. 61–85.

[6] O. C. J. Matas, M. Urban and T. Pajdla.: Robust wide baseline stereo from maximally stable extremal regions. 13th British Machine Vision Conference, 2002, pp. 384–393.

[7] K. Mikolajczyk and C. Schmid.: An affine invariant interest point detector. Computer Vison - Eccv 2002, 2002, pp. 128–142.

[8] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir and L. van Gool.: A comparison of affine region detectors. International Journal of Computer Vision, Vol. 65, 2005, No. 1-2, pp. 43–72.

[9] L. J. V. Gool, T. Moons and D. Ungureanu.: Affine/ photometric invariants for planar intensity patterns. 4th European Conference on Computer Vision, Vol. 1, Springer-Verlag, 1996, pp. 642-651.

[10] F. Mindru, T. Tuytelaars, L. Van Gool and T. Moons.: Moment invariants for recognition under changing viewpoint and illumination. Computer Vision and Image Understanding, 2004, No. 1-3, pp. -3-27.

[11] W. T. Freeman and E. H. Adelson.: The design and use of steerable filters. Ieee Transactions on Pattern Analysis and Machine Intelligence, Vol. 13, 1991, No. 9, pp. 891–906.

[12] B. t. H. R. L.M.T. Florack, J.J Koenderink, and M.A. Viergever.: General intensity transformations and differential invariants. Journal of Mathematical Imaging and Vision, Vol. 4, 1994, No. 2, pp. 171–187.

[13] A. Baumberg.: Reliable feature matching across widely separated views. CVPR 2000: IEEE Conference on Computer Vision and Pattern Recognition, Institute of Electrical and Electronics Engineers Computer Society, Los Alamitos, CA, USA, 2000, pp. 774–781.

[14] F. Schaffalitzky and A. Zisserman.: Multi-view matching for unordered image sets, or How do i organize my holiday snaps? Computer Vison - Eccv 2002, 2002, pp. 414–431.

[15] G. Carneiro and A. D. Jepson.: Multi-scale phase-based local features. 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Institute of Electrical and Electronics Engineers Computer Society, 2003, pp. I/736–I/743.

[16] Y. Ke and R. Sukthankar.: Pca-sift: A more distinctive representation for local image descriptors. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004, Institute of Electrical and Electronics Engineers Computer Society, Piscataway, NJ 08855-1331, United States, 2004, pp. II506-II513.

[17] N. E. Huang, Z. Shen, S. R. Long, M. L. C. Wu, H. H. Shih, Q. N. Zheng, N. C. Yen, C. C. Tung and H. H. Liu.: The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences, Vol. 454, 1998, No. 1971, pp. 903–995.

[18] Z. Liu, H. Wang and S. Peng.: Texture segmentation using directional empirical mode decomposition. 2004 International Conference on Image Processing, ICIP

2004, Institute of Electrical and Electronics Engineers Computer Society, Piscataway, NJ 08855-1331, United States 2004, pp. 279–282.

[19] P. Flandrin, G. Rilling and P. Goncalves.: Empirical mode decomposition as a filter bank. Ieee Signal Processing Letters, Vol. 11, 2004, No. 2, pp. 112–114.

[20] G. Rilling, P. Flandrin and P. Goncalves.: Empirical mode decomposition, fractional gaussian noise and hurst exponent estimation. 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '05, Institute of Electrical and Electronics Engineers Inc., Piscataway, NJ 08855-1331, United States, 2005, pp. IV489–IV492.

[21] J. C. Nunes, Y. Bouaoune, E. Delechelle. Texture analysis based on local analysis of the Bidimensional Empirical Mode Decomposition. Machine Vision and Application, 2005, No. 16, pp, 177–188.

[22] G. Carneiro and A. Jepson. Phase-based local features. In European Conference on Computer Vision, Copenhagen, Denmark, May 2002.

[23] http://www.robots.ox.ac.uk/∼vgg/research/affine/