# Fine Tuning of Quasi Linear Feature Descriptors by the Means of SSD Error Measure

Zoltán Prohászka

Budapest University of Technology and economics,
Dept. of Control Eng. and Information Technology
Magyar Tudosok krt 2, H-1117 Budapest, Hungary
http://www.iit.bme.hu/~prohaszka
e-mail:prohaszka@iit.bme.hu

**Abstract.** This paper focuses on the optimal weighting of the components of a rotation invariant feature vector. This feature descriptor is not expected to be outstanding in performance, it is published here to illustrate the mathematics of the tuning problem. Theoretical tools are used to find proper distance functions. It is assumed, that the resulting error is properly expressed as the sum of squared pixel differences of the corresponding images. This leads to closed formulae for the elements of the weighting matrix. The deductions are intended to be general enough, enabling the application to any linear feature vector. Test results are presented to show differences between uniform and weighted distance measures. Application of the results to the Self Affine Feature Transform (SAFT) is shown briefly.

**Keywords:** Feature-descriptors, sum of squared differences, geometric error, orthogonal weighting functions

## 1 Introduction

Feature descriptors are extensively used in computer vision (CV) and image processing (IP). Many of them is used by the community as a standard tool (SIFT [1], MSER [2], SURF [3], [4]). By photograph registration, features are compared by Euclidean distance in the space of descriptor elements. When a new descriptor is introduced, this distance measure must be carefully investigated and modified. It is not the purpose of the article to overview or to compare different feature descriptors, it focuses rather on the general problem of fine tuning of them, by adjusting the importance of the elements of the feature vector. The author faced this problem during the enhancement of the Self Affine Feature Transform (SAFT) descriptor to enable matching features on photographs [7].

Kronecker product is notated by $\otimes$. $\overline{\mathbf{M}}^{\mathrm{T}}$ will denote Hermitian adjoint of $\mathbf{M}$, and $\overline{z}$ is complex conjugate of $z$.

### 1.1 The Basic Problem

Let there be given the following 2-Dimensional (2-D) IP problem: Let $A$ be a circular investigation window with radius $r_{max}$, whose size and position was determined by a proper pre-processing algorithm. We want to compare the contents of this window to other similarly selected windows, and determine, which one's contents are the most similar to the first window contents. To do so, we determine a feature vector based on the windows' contents. We choose the formulae in such a way that ensures that the posted task can be solved by comparing these feature vectors only. The task will be extended to be invariant against different rotations of different windows. We want to fulfill two requirements:

- The distance of the feature vectors should represent similarity as properly as possible
- This distance function should be invariant against rotations

To achieve simple mathematical formulae, feature vectors linearly depending on pixel values are defined and used. This implies that feature descriptor elements can be interpreted as weighted averages of the image. Thus, each descriptor element has its own weighting function describing which part of the image is represented by itself. The reader must pay attention that the term weighting is used in this text to refer the weighted avarage of pixels and also the scaling of descriptor elements which tunes distance measuring. Moreover, the whole investigation window can have a global weighting function, such as a Gaussian or a constant weighting function.

### 1.2 Basic Formulation of the Investigated Linear Descriptor

Let $I(x, y)$ be the function describing the gray scale image. Let us introduce a general linear descriptor, $h_{k,l} \in \mathbb{C}, k = 0 \dots n, l = 0 \dots m$:

$$h_{k,l} = \int_A w_{h_{k,l}}(r, \phi) I(x, y) \mathrm{d}A \bigg/ \int_A \mathrm{d}A, \tag{1}$$

where $r = x^2 + y^2, \quad \phi = \arg(x + iy), \quad w_{h_{k,0}}(r, \phi) = r^k, \quad w_{h_{k,l\neq 0}}(r, \phi) = 2r^k e^{il\phi}$.

## 2 Rotational Invariance

The rotation-invariant comparison of feature descriptors of this kind can be formalized by introducing parameter $\alpha$, the angle of free rotation. We have to find the optimal $\alpha$ yielding to minimum comparison error.

Let us suppose, that the weighting functions do not change by any rotation of the coordinate frame around the window center, or they can be coupled into complex valued pairs so they change in the following manner due to rotations:

$$w'_{\mathbb{C}}(x, y) = e^{pi\alpha} w_{\mathbb{C}}(x, y), \tag{2}$$

$$w'_{\mathbb{C}} = w'_1 + iw'_2, \qquad w_{\mathbb{C}} = w_1 + iw_2, \qquad p \in \mathbb{Z}^+,$$

(where $w_1$ and $w_2$ are corresponding couples of the pair-vise original weighting functions). In this case the following method can be applied to determine the angle of rotation yielding to minimum squared difference of feature vector elements. Note, that the weighting functions do not need to form an orthogonal system. Even for a given $z$ parameter, there can be more coupled weighting functions (with the same or different radial distribution and swirl). Let us assume, we have $N$ descriptor elements: $v_q, q = 1 \dots N$, which utilize only the above described weighting functions. Let the $N \times N$ diagonal matrix $\mathbf{P}$ describe index $p$, the harmonic frequency of rotational dependence.

If we have two feature vectors $\mathbf{v_A}$ and $\mathbf{v_B}$ coming from image A and B, the squared difference of the features are:

$$e^2_{AB} = |\mathbf{v_A} - \mathbf{v_B}_{rotated}|^2.$$

If image B is rotated by angle $\alpha$, the squared error-distance will be function of $\alpha$:

$$e^2_{AB}(\alpha) = |\mathbf{v_A} - e^{\mathbf{P}i\alpha}\mathbf{v_B}|^2. \tag{3}$$

$$e^2_{AB}(\alpha) = \sum_{q=1}^{N} |v_{Aq}|^2 + |v_{Bq}|^2 - 2\Re(\overline{v_{Aq}} e^{P_{qq}i\alpha} v_{Bq}), \tag{4}$$

This summarization can be grouped according to $P_{qq}$, yielding to :

$$e^2_{AB}(\alpha) = \sum_{p=0}^{p_{max}} \sum_{q:P_{qq}=p} c_q + \Re(d_q e^{pi\alpha}),$$

where

$$c_q = |v_{Aq}|^2 + |v_{Bq}|^2, \qquad d_q = -2\overline{v_{Aq}} v_{Bq}. \tag{5}$$

$$e^2_{AB}(\alpha) = \sum_q c_q + \sum_{p=0}^{p_{max}} \Re(d'_p e^{pi\alpha}), \tag{6}$$

and $d'_p = \sum_{q:P_{qq}=z} d_q$. The minimum of this function is the minimal distance of $\mathbf{v}_A$ and $\mathbf{v}_B$ regarding rotations. This distance function has the advantages that in the case of rotational symmetry, it will have multiple minimum locations.

$\alpha_{opt}$, the optimal angle can be found by brute-force method, $e^2_{AB}$ should be sampled in $o \times p_{max}$ locations, $o \approx 5 \dots 7$, a parabola should be fitted to 3 samples near the minimum, and the minimum of this parabola needs to be located.

In the following let us illustrate this method using descriptor $h_{k,l}$ defined in (1). If $B$ is rotated by $\alpha$ degree, then $h'_{Bk,l} = e^{li\alpha} h_{Bk,l}$. Thus the squared difference is the function of $\alpha$:

$$e^2_{AB}(\alpha) = \sum_{k,l} |h_{Ak,l} - e^{li\alpha} h_{B0,l}|^2 \tag{7}$$

Applying equations (5) and (6), we can see that the squared error is a $m^{th}$ order Fourier series function of $\alpha$.

$$e_{AB}^2 = \sum_{k=0}^{n} \sum_{l=0}^{m} |h_{A k,l}|^2 + |h_{B k,l}|^2 - 2 \cdot \Re(e^{li\alpha} \overline{h_{A 0,l}} h_{B 0,l}). \tag{8}$$

## 3 The Optimal Descriptor Distance Function

After implementing and testing the algorithm above, we faced the question wether uniform weighting of feature vector elements is the optimal choice, or we can find better distance functions by weighting the linear combination of the (complex) elements ($\overline{\mathbf{h}}^{\mathrm{T}}\mathbf{h}$ vs. $\overline{\mathbf{h}}^{\mathrm{T}}\mathbf{T}^{-1}\mathbf{h}$).

In this point we tried to find formulae, which ensure that the algebraic error function will be very similar to the geometric error. This technique is advised throughout [5] and is also discussed in [6]. However, the IP problems appearing in [5] suggests that the fulfillment of these requirements needs computationally expensive itereations. Fortunately, for many preprocessing problem (together with the problem discussed here) the requirements above can be satisfied by relatively fast methods and formulation.

According to the above, a more precise interpretation of 'windows with similar contents' was given, namely we are seeking for windows, whose integrated squared pixel differences is minimal (after rotating the second one with the optimal angle). Obviously, the information content of these windows is much higher then the degrees of freedom of the feature descriptors. Therefore, there will be a lot of information on the images (in the null-space of the weighting functions) which will have no effect on our algorithm.

Since the introduced weighting functions are the direct products of radial and tangential weighting functions $w_{h k,l} = w_{rad k} w_{tan l}$, the problem can be formulated with Kronecker products. If we collect the previous weighting functions into vectors, we get:

$$\mathbf{w} = \mathbf{w}_{rad} \otimes \mathbf{w}_{tan}, \quad \mathbf{w} :_{(n+1)(m+1)\times 1}, \mathbf{w}_{rad} :_{(n+1)\times 1}, \mathbf{w}_{tan} :_{(m+1)\times 1} . \tag{9}$$

Notice, that this formulation determines the order of $h_{k,l}$ descriptor elements in the feature vector $\mathbf{h}$. In other situations, $m$ can be chosen to 0, and the formulae can be simplified using $\mathbf{A} \otimes 1 = \mathbf{A}$.

If all radial weight functions are orthogonal and have the same strengths, then $\overline{\mathbf{h}}^{\mathrm{T}}\mathbf{h}$ would be a good error measure. The problem comes if the dot (scalar) product of different weighting functions is non-zero.

Let $\mathbf{T}_{rad}$: $(n+1) \times (n+1)$ matrix contain these scalar products:

$$T_{rad k_1,k_2} = \int_A w_{rad k_1}(x,y) w_{rad k_2}(x,y) w_W(x,y) \mathrm{d}A, \tag{10}$$

where $w_W(x,y)$ is the weighting of the window. Applying (10) to tangential functions and using (9) leads to:

$$\mathbf{T} = \mathbf{T}_{rad} \otimes \mathbf{T}_{tan}. \tag{11}$$

These matrices can be calculated either by numerically summarizing pixels, or analytically also. In our application $w_{radk}(r,\phi) = r^k, w_W(r) = 1$, thus $T_{radk_1,k_2} = 1/(k_1+k_2+2)r_{max}^{(k_1+k_2+2)}$. If we transform $\mathbf{w}_{rad}$, the vector of weighting functions by

$$\mathbf{w}'_{rad} = \mathbf{U}_{rad}\mathbf{w}_{rad}, \tag{12}$$

where $\mathbf{U}_{rad} = \mathbf{T}_{rad}^{-0.5}$, and calculate similarly $\mathbf{U}_{tan}$ and use $\mathbf{U} = \mathbf{U}_{rad} \otimes \mathbf{U}_{tan}$, then we receive orthogonal weighting functions. Descriptor elements must be combined by the same way (due to linearity):

$$\mathbf{h}'_{rad} = \mathbf{U}_{rad}\mathbf{h}_{rad}, \qquad \mathbf{h}'_{tan} = \mathbf{U}_{tan}\mathbf{h}_{tan} \quad \Rightarrow \quad \mathbf{h}' = \mathbf{U}\mathbf{h}. \tag{13}$$

It is evident, that $\mathbf{U}_{rad}$ (or $\mathbf{U}_{tan}$ or $\mathbf{U}$) can be multiplied from left by any rotation matrix $\mathbf{Q}_{rad}^{\mathrm{T}}$, so if someone would like to have 'standard' radial weight functions, he/she can use the following choices got by QR factorization:

$$\mathbf{U}'_{rad} = \mathbf{R}_{rad}, \quad \mathbf{Q}_{rad}\mathbf{R}_{rad} = \mathbf{U}_{rad}$$
$$\hat{\mathbf{U}}'_{rad} = \hat{\mathbf{R}}_{rad}, \quad \hat{\mathbf{Q}}_{rad} flip_{\leftrightarrow}(\hat{\mathbf{R}}_{rad}) = flip_{\leftrightarrow}(\mathbf{U}_{rad}),$$

where $flip_{\leftrightarrow}(\ )$ flips matrices horizontally. This is a common technique in the determination of Legendre polynomials, thus this paper is strongly connected to them. (If lower order weighting functions appear in the bottom of $\mathbf{w}_{rad}$, use $\mathbf{U}'_{rad}$, otherwise use $\hat{\mathbf{U}}'_{rad}$.)

### 3.1 Combining rotational invariance and optimal weighting

Now we can utilize, that for the $h_{k,l}$ descriptor $\mathbf{T}_{tan} = \mathbf{U}_{tan} = \mathbf{I}$, thus if we organize the elements of $\mathbf{h}$ into a complex matrix $\mathbf{H}_{k,l} = h_{k,l}$ having $k+1$ for vertical index and $l+1$ for horizontal index, then only the same $\mathbf{U}_{rad}$ weighting matrix needs to be applied to every column:

$$[SSD_{min}, \alpha_{opt}] = optimalphase(\mathbf{U}\mathbf{H}_A, \mathbf{U}\mathbf{H}_B).$$

where *optimalphase*( ) denotes the algorithm described in section 2 in equations (5) to (8).

## 4 Test Results

After implementing the algorithms above, we performed the following tests.

200 test images were randomly generated, which have rotational symmetry in some cases (depending on random events). Each image's 2nd, 1st and 0th moment is normalized before processing. In Fig. 1 and Fig. 2, the same, randomly selected eight images are shown without weighting and with optimal weighting respectively. The closest pair found by the algorithm, and the difference of reproduced images are also shown. Image center is underrepresented in the unweighted case.

We have to highlight again, that the weighting matrix is also affected by the global weighting function applied to every window of interest. In the tests above, such place-dependent weighting was not used.

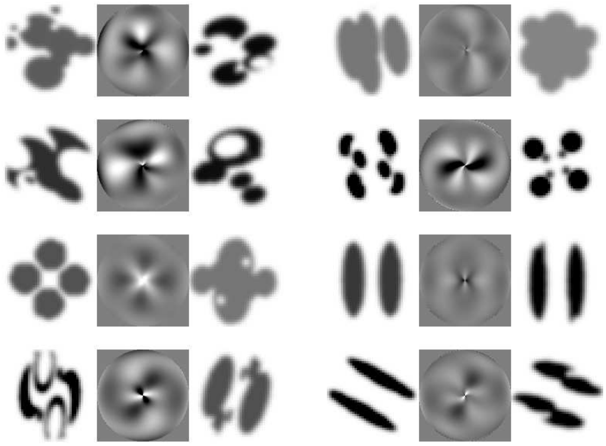

**Fig. 1.** Image pairing without weighting matrix. Left column: Normalized Images ($A$); Right column: The closest neighbor found by $opt(\overline{\mathbf{h}}^{\mathrm{T}}\mathbf{h})$, normalized and rotated to best alignment $Rot(B, \alpha_{opt})$; Middle column: Difference of the reproduced images. $Reproduce(Extract(A - Rot(\mathrm{B}, \alpha_{opt}))) + 0.5$

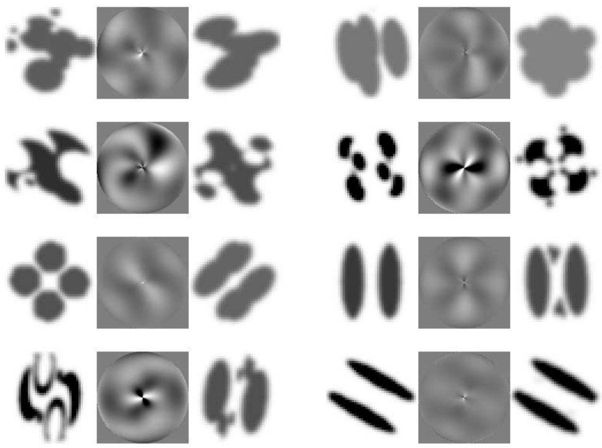

**Fig. 2.** Image pairing with weighting matrix enabled. Notice better pairing, better alignment, less error especially in the center.

## 5 Fine Tuning of SAFT

If the SAFT feature descriptor is used for photograph registration, then the careful transformation of the 18-dimensional feature vector is needed to ensure that feature distance will correlate to the SSD pixel error of the contents of compared investigation windows. The results described above can be applied with some modification while the SAFT feature is not linear to pixel values.

A comprehensive introduction to the SAFT method can be found in [7]. It focuses on the analysis of drawings and photographs of engineered objects. However, the formulation given applies here also. The main difference is that we have to use the Kronecker-product in reversed order, than used in [7]. Let us denote this permutated vectors and matrices with $\hat{}$. Then, $\hat{\mathbf{M}}$ can be divided into three different $3 \times 3$ symmetric matrices:

$$\hat{\mathbf{M}} = \begin{bmatrix} \mathbf{G}_{xx} & \mathbf{G}_{xy} \\ \mathbf{G}_{xy} & \mathbf{G}_{yy} \end{bmatrix}. \tag{14}$$

It is advantageous to take the linear combination of them:

$$\mathbf{G}_K = \mathbf{G}_{xx} + \mathbf{G}_{yy}, \quad \mathbf{G}_C = \mathbf{G}_{xx} - \mathbf{G}_{yy}, \quad \mathbf{G}_S = 2\mathbf{G}_{xy}. \tag{15}$$

If we calculate the image of squared gradient magnitudes $I_G$,

$$I_G(x, y) = |\nabla I(x, y)|^2 \tag{16}$$

then the elements of $\mathbf{G}_K$ are linear combination of $I_G$ pixel values.

$$\mathbf{G}_K = \int_A I_G \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \begin{bmatrix} x \; y \; 1 \end{bmatrix} \mathrm{d}A \tag{17}$$

Therefore, the optimal weighting of the $\mathbf{G}_K$ elements can be deduced by the means of the algorithm shown in section (3):

$$\mathbf{G}_K = \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} a \; b \; c \; d \; e \; f \end{bmatrix}^{\mathrm{T}}, \quad \mathbf{v}' = \mathbf{U}'\mathbf{v}. \tag{18}$$

If the investigation window's radius is 2 (according to [7]) then

$$\mathbf{T} = \frac{1}{3} \begin{bmatrix} 6 & 0 & 0 & 2 & 0 & 3 \\ 0 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 12 & 0 & 0 & 0 \\ 2 & 0 & 0 & 6 & 0 & 3 \\ 0 & 0 & 0 & 0 & 12 & 0 \\ 3 & 0 & 0 & 3 & 0 & 3 \end{bmatrix}, \quad \mathbf{U}' = \frac{1}{4} \begin{bmatrix} -2\sqrt{3} & 0 & 0 & -2\sqrt{3} & 0 & 4\sqrt{3} \\ 0 & 0 & 0 & 0 & 0 & 4 \\ \sqrt{6} & 0 & 0 & -\sqrt{6} & 0 & 0 \\ 0 & \sqrt{6} & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \end{bmatrix}, \tag{19}$$

where $\mathbf{U} = \mathbf{T}^{-0.5}, \mathbf{U}' = \mathbf{QU}$ and $\mathbf{QQ}^{\mathrm{T}} = \mathbf{I}$ is a chosen rotation. The elements of $\mathbf{G}_K$ lie in a subspace of the elements of $\tilde{\mathbf{M}}$. In this article we do not discuss the weighting of other components ($\mathbf{G}_C$ and $\mathbf{G}_S$). It will be shown in a longer article describing many aspects and test results of SAFT used as a photograph matcher, which article will refer to the formulae defined here extensively.

## 6 Conclusion

This article discussed the problem of finding proper distance functions for linear and quasi-linear feature descriptors. A simple descriptor was introduced to illustrate the mathematical background. Rotational Invariance of a specialized distance function was shown. An optimal weighting matrix is presented, which ensures that the algebraic error is similar to the geometric error. These results were described generally, indicating which conditions are required to apply these results to other feature descriptors. The results were applied in the tuning of the SAFT feature descriptor in an illustrative way.

**Acknowledgments.** The research in this article was supported by the Hungarian National Research Program under grant No. OTKA K 71762.

## References

1. Lowe, D. Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision, pp ? (2003)
2. Matas, J. Chum, O., Urban, M., and Pajdla, T. Robust wide-baseline stereo from maximally stable extremal regions. In Proceedings of the British Machine Vision Conference, Cardiff, UK, pp. 384-393. (2002)
3. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.:SURF: Speeded Up Robust Features, Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346–359, (2008)
4. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L.: A Comparison of Affine Region Detectors, Int. J. Comput. Vision, Vol. 65, pp.43-72. (2005)
5. Hartley, R., Zissermann A.: Multiple View Geometry in Computer Vision. Cambridge University Press, (2003)
6. Faugeras, O., Luong, Q.-T.:The Geometry of Multiple Images, MIT Press. (2001)
7. Prohaszka, Z.: Affine Invariant Features from Self-Flow In: Proceedings of the RAAD 2008 17th International Workshop on Robotics in Alpe-Adria-Danube Region September 15-17, 2008, Ancona, Italy. pp. 100, ,Ancona (2008). `http://mycite.omikk.bme.hu/doc/40788.pdf`