Connection of the Harris Detector-Based Affine Adaptation and the Self Affine Feature Transform

Zoltan Prohaszka[1]

Department of Control Engineering and Information Technology,
Budapest University of Technology
prohaszka@iit.bme.hu

Abstract. The Self Affine Feature Transform (SAFT) can both analyze geometric image details and match similar regions of photographs. This article presents connections between the Harris detector based Affine Adaptation (Harris-Affine) and the SAFT descriptor. Based on these, a novel evaluation of the matrix of the SAFT descriptor is presented. Possible applications of the shown relations during affine shape adaptation are also highlighted.

1 Introduction

This article presents interesting relations between the Self Affine Feature Transform (SAFT) and the Harris detector based Affine Shape Adaptation (Harris-Affine), utilizing special properties of the Gaussian window. It will be shown, that the linearized dependence of the Harris matrix from the parameters of the used integration kernel is encoded in the SAFT matrix. For moderate affine transformations, it can be used for an alternative evaluation of the Harris-Affine iterative method. It is expected to improve the speed of the method. The described relations are used to present a new evaluation of the SAFT descriptor.

The article is organized as follows: A short introduction is given from related research work. Then the two methods related to the article (SAFT and Harris-Affine) are shown with more details. The equivalence of the SAFT descriptor and the linearized dependence of the Harris matrix for infinite Gaussian windows are shown. The last section of the article presents an alternative evaluation of the SAFT matrix together with test results.

1.1 Related Work

The Self Affine Feature Transform (SAFT) was introduced by Prohaszka in 2008 [1]. It can extract geometric information from 2D grayscale images invariantly from regular affine transformations. It can be used also to compress image detail information into a 54-dimensional feature vector, which consist of 3 18-DoF SAFT descriptor evaluated using different image frequencies. This enables it to

be used for photograph registration also, which is shown in [2] Its purely analytical, continuous formulation enables a wide variety of post-processing algorithms to be used on the descriptor matrix [3], and easy implementation on GPUs. These properties make SAFT be different from standard feature descriptors, which are mainly defined by algorithms, often having branches and discrete tables. Wide spread feature descriptors use combined domain and range histograms of gradients: SIFT [4], GLOH [5]. SAFT relates to several moment extracting methods, for example RGB moments used in [6], however SAFT uses combined domain and range moments of gradients.

Affine shape adaptation is used to detect regions of an image which are stable against affine transformation. This task was solved for similarity transformations by the Scale-Space theory [7, 8], its extension gives scale invariance to the Scale Invariant Feature Transform [4]. Lindeberg and Garding gave affine extension of the Scale-Space algorithm and proposed an iterative algorithm to solve the above described task [9]. Mikolajczik and Schmid made improvements on the method [10] which is referred as Harris-Affine region detector in the comparative article [11]. MSER is a completely different affine invariant region detector [12]. It analizes image intensity as a height-map. Islands with steep-shores are detected when the height map is flooded with water.

2 The SAFT Detector

This section briefly summarizes previous results about the SAFT detector, which are essential to understand this article. More details can be found in [1] and [2].

2.1 Nomenclature

To distinguish from scalar variables, bold lower-case letters are used for vectors ($\mathbf{v}$) and bold capitals for matrices ($\mathbf{M}$). Vectors are column vectors by default, row vectors appear as transposed columns. $\otimes$ refers Kronecker product of matrices. Its precedence is between $*$ and $+$. Readers not familiar with $\otimes$ need only a very short reading about it. Homogeneous quantities are denoted by subscript $_\mathsf{H}$, to aid differentiating between homogeneous and standard representations. Projective equivalence of homogeneous representations is notated by $\cong$.

Let us consider the function $\mathbf{f}(\mathbf{x})$ over the domain S, $\mathbf{x} \in S$ (both $\mathbf{f}$ and $\mathbf{x}$ can be column vectors or scalars).

The homogeneous 2^{nd} range-, domain- or range-domain-moment of $\mathbf{f}$ is defined by:

$$\int \begin{bmatrix} \mathbf{v}(\mathbf{x}) \\ 1 \end{bmatrix} \begin{bmatrix} \mathbf{v}(\mathbf{x})^\mathsf{T} & 1 \end{bmatrix} \otimes h(\mathbf{x}) \mathrm{d}\mathbf{x}, \tag{1}$$

where $\{\mathbf{v}(\mathbf{x}); h(\mathbf{x})\}$ can be one of the following: $\{[\mathbf{f}(\mathbf{x})]; 1\}$, $\{[\mathbf{x}]; \mathbf{f}(\mathbf{x})\}$ or $\{[\mathbf{f}(x)^\mathsf{T}\mathbf{x}^\mathsf{T}]^\mathsf{T}; 1\}$ respectively.

Notice that the Kronecker product ($\otimes$) usually simplifies to multiplication by scalar, except for domain-moments if $\mathbf{f}$ is vector valued. Range-moments represent function value histograms, while domain-moments represent the distribution of functions on the input space (consider a PDF).

The image $I(x, y)$ is integrated according to the weighting function $w_A(x, y)$ on the domain A. Image gradients are in vector $\mathbf{g}(x, y) = [\frac{\partial I}{\partial x} \ \frac{\partial I}{\partial y}]^\mathsf{T}$. The : (colon) operator appearing in matrix subscripts denotes MATLAB®, OCTAVE style multiple indexing.

The size of a round Gaussian filter is described by σ^2, which is equal to the diagonal elements of its auto-correlation matrix Σ. The quadratic-mean radius of such a filter is $r = \sqrt{N}\sigma$, if the filter is N-dimensional.

2.2 2D Affine Flows

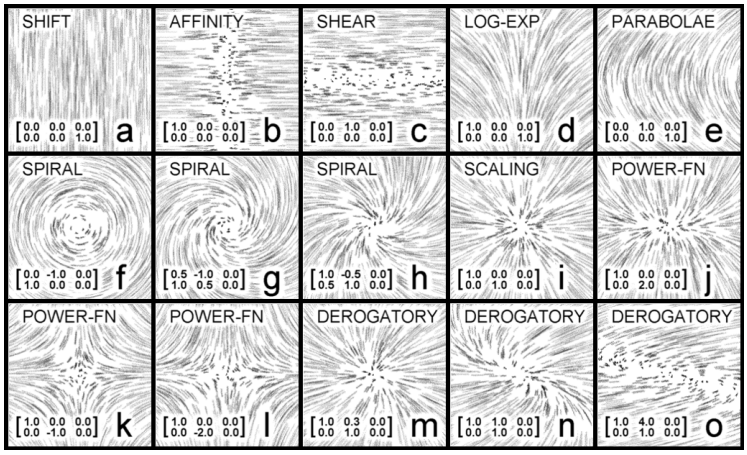

Fig. 1. Typical affine flows, and corresponding $\mathbf{Q}$ matrices.

Affine flows (see Figure 1 for planar examples) in any dimension can be described as local velocity $\mathbf{v}$ is a linear function of homogeneous position $\mathbf{p}_\mathsf{H}$:

$$\mathbf{v} = \mathbf{Q}\mathbf{p}_\mathsf{H} \ , \tag{2}$$

where $\mathbf{Q}$ is a 2×3 matrix. The elements of $\mathbf{Q}$ are the linear parameters of the flow. They can be collected into the parameter vector $\mathbf{q}$ in column major order:

$$\mathbf{v} = \mathbf{L}_2(\mathbf{p}_\mathsf{H})\mathbf{q}, \ \mathbf{L}_2(\mathbf{p}_\mathsf{H}) = \mathbf{p}_\mathsf{H}^\mathsf{T} \otimes \mathbf{I}_{2\times 2} \ , \tag{3}$$

$$\mathbf{L}_2 = \begin{bmatrix} x & 0 & y & 0 & 1 & 0 \\ 0 & x & 0 & y & 0 & 1 \end{bmatrix} . \tag{4}$$

The properties of affine flows are preserved during regular affine transformations.

Affine flows of small strength are identical to infinitesimal affine transformations.

2.3 Image Invariance against Affine Flows

The affine Lucas-Kanade (L-K) detector estimates the optimal affine transformation between two images I and I' [13]. If this affine registration succeeds, the dependence of the registration error against small disturbances of the optimal transformation will be the quadratic function of affine flow parameters. This quadratic function is hold by the local working variables of the L-K detector, and can be managed to be returned also. The investigated region is measured at sample locations which are indexed by $k \in \mathbb{N}$. Each measurement has a local error e_k, which is a linear function of flow parameters. The square of these errors are summed symbolically resulting the quadratic error function:

$$e_k = \mathbf{g}_k^{\mathsf{T}} \mathbf{v}_k - (I(\mathbf{p}_k) - I'(\mathbf{p}_k)),\ e^2(\mathbf{q}) = \sum_k e_k^2(\mathbf{q}) \ , \tag{5}$$

$$e^2(\mathbf{q}) = \begin{bmatrix} \mathbf{q}^{\mathsf{T}} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{M} & \mathbf{n} \\ \mathbf{n}^{\mathsf{T}} & o \end{bmatrix} \begin{bmatrix} \mathbf{q} \\ 1 \end{bmatrix},\ \nabla e^2 = 2\mathbf{q}^{\mathsf{T}}\mathbf{M} + 2\mathbf{n}^{\mathsf{T}} \Rightarrow \mathbf{q}_{opt} = -\mathbf{M}^{-1}\mathbf{n} \ , \tag{6}$$

$$e(\mathbf{q})^2 - e_{opt}^2 = \Delta\mathbf{q}^{\mathsf{T}}\mathbf{M}\Delta\mathbf{q},\ \Delta\mathbf{q} = \mathbf{q} - \mathbf{q}_{opt}. \tag{7}$$

When the two images are identical ($I = I'$), this quadratic function describes the sensitivity of this single image against infinitesimal affine transformations. $\mathbf{M}$, the quadratic matrix of this error function is called the Self-Affine Feature Transform (SAFT):

$$\mathbf{M} = \sum_k \mathbf{M}_k,\ \mathbf{M}_k = \mathbf{L}_k^{\mathsf{T}}\mathbf{C}_k\mathbf{L}_k = \mathbf{p}_{\mathsf{H}k}\mathbf{p}_{\mathsf{H}k}^{\mathsf{T}} \otimes \mathbf{g}_k\mathbf{g}_k^{\mathsf{T}} = \begin{bmatrix} x^2\mathbf{C}_k & xy\mathbf{C}_k & x\mathbf{C}_k \\ xy\mathbf{C}_k & y^2\mathbf{C}_k & y\mathbf{C}_k \\ x\mathbf{C}_k & y\mathbf{C}_k & C_k \end{bmatrix} \ , \tag{8}$$

where $\mathbf{C}_k = \mathbf{g}_k\mathbf{g}_k^{\mathsf{T}}$. Due to this summation, $\mathbf{M}$ preserves positive semi-definite and block-symmetric properties of $\mathbf{M}_k$, therefore it has 18 independent elements.

$\mathbf{M}$ can be partitioned according to the following:

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^{\mathsf{T}} & \mathbf{C} \end{bmatrix} \ , \tag{9}$$

where the 2×2 matrix $\mathbf{C} = \sum_k \mathbf{g}_k\mathbf{g}_k^{\mathsf{T}}$ is the matrix of the 2D Harris corner detector, also known as the second moment matrix (of gradients).

For example, images of concentric circles, polygon vertices and lines are invariant against rotation, scaling around the vertex and against translation respectively. See Figure 2. This can be expressed by algebra as the parameter vector of these flows is in the null-space of the corresponding $\mathbf{M}$ matrix. See [1] for more details, where several methods are given. Figure 3 demonstrates localization of mean circle on images and localization of scaling center.

Regular and degenerate conics can be identified in closed formulation, Figure 4 shows cases when $\mathrm{rank}(\mathbf{M}) = 4$.

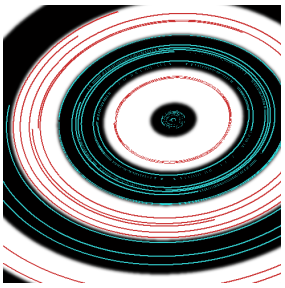

Fig. 2. Concentric ellipses and streamlines of their invariant flow.

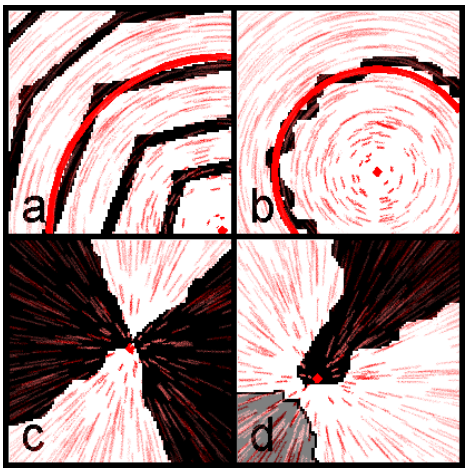

Fig. 3. SAFT can be used to find optimal center of rotation (a,b) or scaling (c,d) in closed form.

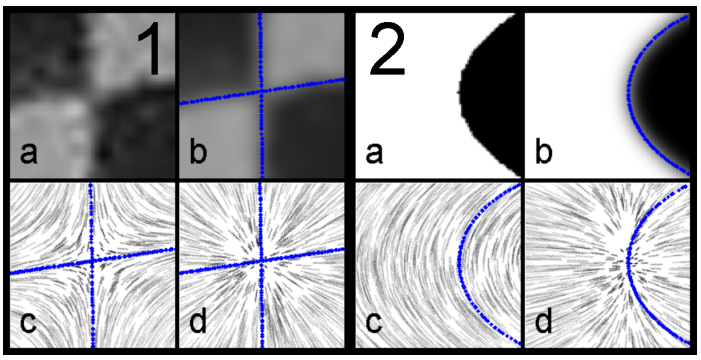

Fig. 4. Conics resulting rank 4 **M** matrices are precisely reconstructed.

2.4 Registration of photographs by SAFT

Main steps of the registration of two photographs are:

- Localization of keyframes
- Compressing visual contents of keyframes into descriptor vectors
- Selecting matching candidates based on descriptor vector distances
- Robust determination of transformation model by validating matching candidates based on their geometric arrangement.

SAFT offers an alternative solution for descriptor extraction. Multiple band approach is used, the 18-DoF SAFT matrix $\mathbf{M}$ has to be evaluated using different frequencies of the image signal. 3 bands are considered to be enough. Preliminary parametrization of the 3 bands described in [2] has a performance gap relative to the SIFT descriptor, which was preliminary expected based on the difference of descriptor vector size (54 for SAFT and 128 for SIFT). Further optimization resulted parameter-triplets which can perform slightly better than SIFT during large disparity binocular registration. The involved evaluation of the SAFT matrix uses Finite Impulse Response (FIR) Gaussian windowing.

3 Harris-based Affine Shape Adaptation

Lindeberg described a possible iterative algorithm to localize regions which are fixed for affine transformations, based on the second moment matrix of gradients ($\mathbf{C}$) [9]. Mikolajczyk made some improvements on the method [10] and give detailed performance tests in [11]. The basic idea behind the method is that the images are supposed to contain regions which are affinely distorted between views. Results of Scale-Space theory are generalized to the affine case: No preliminary information can be used about the size (and shape) of the differentiating and integrating kernels used during the processing of the image. Many sizes (and shapes) has to be used to extract local image features. Only the output of those filters can be accepted, whose kernel sizes (and shapes) are validated by output results. This validation for the Harris-Affine method requests that

- the shape of the differentiating and integrating kernels must be the same.
- The second moment matrix $\mathbf{C}$ (matrix of the Harris detector) must have a shape according to the used kernels.
- the size and position of the integrating kernel must yield to local maximum of a scalar which is a function of normalized second moment matrix $\sigma_{dif}^2\mathbf{C}$, for example trace($\sigma_{dif}^2\mathbf{C}$).
- the ratio of the size of the integration and differentiating kernels ($\sigma_{int}/\sigma_{dif}$) can be fixed, in this case this ratio is a parameter of the method.

When an elliptic region of the images satisfies these conditions, then any (non-singular) affine transformed view of it satisfies the conditions, as long as sampling effects does not play role (the lower σ parameter of the differentiating kernel is larger than $1 \ldots 2$ pixels).

The task is to find this kind of fixed-points of affine transformations of the image quickly. Highlights of the proposed iterative methods are:

1. initiating kernel shapes to circular
2. finding local maximum in scale and position
3. evaluating $\mathbf{C}$ using actual kernels
4. updating kernel shapes
5. go back to 2 until convergence occurs or fails.

The most intensive task during this iteration is to evaluate the Harris matrix $\mathbf{C}$ using the same differentiating kernel but slightly displaced integration kernels while seeking maximum on the plane. This is repeated for at least 3 different value of scale.

4 Equivalence

This section proves, that the SAFT descriptor is equivalent to the linearized model of the Harris-Affine method, if distortions of the differentiating kernel are not taken into account.

If the SAFT matrix $\mathbf{M}$ is evaluated on an origin-centered unit deviation Gaussian window, then its components can be described as integrating matrix $\mathbf{C}$ with weighting functions $x^p y^q \frac{1}{2\pi} e^{-[x\ \ y][x\ \ y]^\mathsf{T}/2}$. Due to the special properties of the Gaussian function, these weighting functions corresponds to planar derivatives of the Gaussian window of orders p and q. Moreover, these functions also correspond to the infinitesimal change of the weighting function with respect to its parameters like shift, scaling and affinity in different directions:

$$\mathbf{C}(\mathbf{T}_\mathsf{H}) = \frac{\int \mathbf{g}\mathbf{g}^\mathsf{T} e^{-\mathbf{p}_\mathsf{H}^\mathsf{T}\mathbf{T}_\mathsf{H}^\mathsf{T} D_{iag}([1\ 1\ 0])\mathbf{T}_\mathsf{H}\mathbf{p}_\mathsf{H}/2}\mathrm{d}A}{\int 1 \cdot e^{-\mathbf{p}_\mathsf{H}^\mathsf{T}\mathbf{T}_\mathsf{H}^\mathsf{T} D_{iag}([1\ 1\ 0])\mathbf{T}_\mathsf{H}\mathbf{p}_\mathsf{H}/2}\mathrm{d}A}, \tag{10}$$

where $\mathbf{p}_\mathsf{H} = [x\ \ y\ \ 1]^\mathsf{T}$. The shape and position of the integration kernel are described by

$$\mathbf{T}_\mathsf{H} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0\ 0 & 1 \end{bmatrix}.$$

The points of the main ellipse of the integration kernel transforms to the unit circle by multiplying their column vector representation from left by $\mathbf{T}_\mathsf{H}$. Then (10) simplifies to:

$$\mathbf{C}(\mathbf{T}_\mathsf{H}) = \frac{|\mathbf{R}^\mathsf{T}\mathbf{R}|}{2 \cdot \pi} \int \mathbf{g}\mathbf{g}^\mathsf{T} e^{-\mathbf{p}_\mathsf{H}^\mathsf{T}\mathbf{T}_\mathsf{H}^\mathsf{T} D_{iag}([1\ 1\ 0])\mathbf{T}_\mathsf{H}\mathbf{p}_\mathsf{H}/2}\mathrm{d}A. \tag{11}$$

The dependence of $\mathbf{C}$ from $\mathbf{t} = [t_x\ \ t_y]^\mathsf{T}$ (given $\mathbf{R} = \mathbf{I}_{2\times 2}$) is the following:

$$\partial\mathbf{C}/\partial t_x = \frac{1}{2\pi} \int -\mathbf{g}\mathbf{g}^\mathsf{T} x e^{-\mathbf{p}_\mathsf{H}^\mathsf{T}\mathbf{I}_{2\times 2}\mathbf{p}_\mathsf{H}/2}\mathrm{d}A = -2x\mathbf{C}(\mathbf{I}_{3\times 3})/2 \tag{12}$$

$$\partial\mathbf{C}/\partial t_y = \frac{1}{2\pi} \int -\mathbf{g}\mathbf{g}^\mathsf{T} y e^{-\mathbf{p}_\mathsf{H}^\mathsf{T}\mathbf{I}_{2\times 2}\mathbf{p}_\mathsf{H})/2}\mathrm{d}A = -2y\mathbf{C}(\mathbf{I}_{3\times 3})/2,$$

If we introduce scalars k, c, s to encode changes of the shape of the applied Gaussian bell according to

$$\mathbf{R}^{\mathsf{T}}\mathbf{R} = \mathbf{I}_{2\times 2} + \begin{bmatrix} k+c & s \\ s & k-c \end{bmatrix}, \tag{13}$$

then the derivatives of $\mathbf{C}$ with respect to these parameters evaluated at $\mathbf{t} = [0\ \ 0]^{\mathsf{T}}$ and $\mathbf{R} = \mathbf{I}_{2\times 2}$ are:

$$\partial\mathbf{C}/\partial k = \frac{1}{2\pi}\int \mathbf{g}\mathbf{g}^{\mathsf{T}}(2-(x^2+y^2)/2)e^{-\mathbf{p}_{\mathsf{H}}^{\mathsf{T}}\mathbf{I}_{2\times 2}\mathbf{p}_{\mathsf{H}}/2}\mathrm{d}A = (4-(x^2+y^2))\mathbf{C}(\mathbf{I}_{3\times 3})/2, \tag{14}$$

$$\partial\mathbf{C}/\partial c = \frac{1}{2\pi}\int -\mathbf{g}\mathbf{g}^{\mathsf{T}}(x^2-y^2)/2e^{-\mathbf{p}_{\mathsf{H}}^{\mathsf{T}}\mathbf{I}_{2\times 2}\mathbf{p}_{\mathsf{H}}/2}\mathrm{d}A = -(x^2-y^2)\mathbf{C}(\mathbf{I}_{3\times 3})/2, \tag{15}$$

$$\partial\mathbf{C}/\partial s = \frac{1}{2\pi}\int -\mathbf{g}\mathbf{g}^{\mathsf{T}}xye^{-\mathbf{p}_{\mathsf{H}}^{\mathsf{T}}\mathbf{I}_{2\times 2}\mathbf{p}_{\mathsf{H}}/2}\mathrm{d}A = -2xy\mathbf{C}(\mathbf{I}_{3\times 3})/2. \tag{16}$$

Therefore, the SAFT matrix of infinite Gaussian windows describes the value and dependence of the Harris matrix $\mathbf{C}$ on the parameters of the infinite Gaussian integration kernel, supposed that the same coordinate frame is used during the integrations.

Combining these formulae with the coordinate transformation rules of SAFT, we obtain results for the dependence of the Harris matrix $\mathbf{C}$ on image distortions, however, the distortion of the differentiating kernel is not taken into account. Any usage of SAFT (either information extraction or matching) has some kind of robustness which gives (different levels of) robustness of computation results against anisotropic distortions of the differentiating kernel relative to image contents.

4.1 Shape Adaptation

This subsection describes, how the above described equivalences can be utilized during affine shape application. Standard shape adaptation solutions evaluate $\mathbf{C}$ using a kernel shape, then overwrite the applied shape corresponding to the elements of $\mathbf{C}$. Depending on the window's vicinity, this step in shape parameters c, s can be considerably smaller or higher than the optimal one. (Consider the shape of the letter 'H', and a window which contains only the vertical line of it.) Therefore, it is advantageous to restrict shape change between image integrations. This restriction can be that the transformation between successive kernel shapes can be composed of a rotation and an affinity of a ratio not higher than $2^{\pm 1/2}$.

Let us collect the independent elements of matrix $\mathbf{C}_*$ (with any index) into vector $\mathbf{c}_* = [C_{*1,1}\ \ C_{*1,2}\ \ C_{*2,2}]^{\mathsf{T}}$. In the following equation

$$\begin{bmatrix} g_x^2 \\ g_x g_y \\ g_y^2 \end{bmatrix} = \mathbf{\Gamma}\begin{bmatrix} c \\ s \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{c}_c\ \mathbf{c}_s\ \mathbf{c}_0 \end{bmatrix}\begin{bmatrix} c \\ s \\ 1 \end{bmatrix} \approx \mathbf{c}(c,s) \tag{17}$$

$\mathbf{\Gamma}$ contains the linearized dependence of $\mathbf{C}$ from shape changes described by c, s, $\mathbf{c}_0 = \mathbf{c}(\mathbf{I}_{3\times3})$, $\mathbf{c}_s = \partial\mathbf{c}/\partial s(\mathbf{I}_{3\times3})$ and $\mathbf{c}_c = \partial\mathbf{c}/\partial c(\mathbf{I}_{3\times3})$.

The shape described by $\mathbf{c}$ must coincide with the shape of the integration window (described by $[1+c \;\; 2s \;\; 1-c]$), yielding the eigenvalue problem:

$$\mathbf{\Gamma}\begin{bmatrix} c \\ s \\ 1 \end{bmatrix} = \lambda \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} c \\ s \\ 1 \end{bmatrix} \tag{18}$$

The exact solution could be obtained by calculating the eigenvectors of

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ -1 & 0 & 1 \end{bmatrix}^{-1} \mathbf{\Gamma}.$$

Since the model utilized is already a linear approximation, a fast approximation of the eigenvectors is appropriate.

To do so, an iteration has to be started by evaluating the SAFT matrix using the current kernel shape. The obtained matrix has to be coordinate transformed to the frame representing the kernel shape used. Then $c = 0$ and $s = 0$ has to be set. The matrix $\mathbf{\Gamma}$ must be normalized in each step as $\mathbf{\Gamma}[c \;\; s \;\; 1]^\mathsf{T}$ has unit length. Then, the two sides of (18) can be solved for minimum squared difference. When the shape change during this iteration exceeds the above described limit, new SAFT measurements have to be carried out on the image. This modification can be performed by integrating only 9 required components of $\mathbf{M}$ instead of 18. If the image contains a fixed-point exactly lying under the integration kernel of SAFT, then this is signaled by

$$M_{5,6} = 0 \;\; \textit{and} \;\; M_{5,5} = M_{6,6}. \tag{19}$$

According to [1], the SAFT matrices of regular shapes satisfy these equations.

Local maxima of trace($\mathbf{C}$) *on the plane:* The trace of $\mathbf{C}$ has a maximum with respect to translations, if $\partial\text{trace}(\mathbf{C})/\partial x = 0$ and $\partial\text{trace}(\mathbf{C})/\partial y = 0$. These conditions are identical to

$$M_{1,5} + M_{2,6} = 0, \;\; M_{3,5} + M_{4,6} = 0. \tag{20}$$

Moreover, if the trace is normalized with respect to different scales as the effect of the differentiating filter not being taken into account, only the area growth of the integration kernel is observed, then a maximum of normalized trace can exist if $\partial\text{trace}(\mathbf{C})/\partial k = 0$. This corresponds to

$$\sum_{l=1}^{4} M_{l,l} = 4(M_{5,5} + M_{6,6}). \tag{21}$$

The SAFT matrix of regular shapes in the origin (according to [1]) satisfies these equations if the coordinate unit is chosen appropriate.

5 Alternative evaluation of SAFT

Let us investigate the 1D function:

$$f(x) = xG(x, \sigma),$$

where $G(x, \sigma) = 1/(\sqrt{2\pi\sigma^2})e^{-(x/\sigma)^2/2}$ is the 1D Gaussian function. $f(x)$ can be evaluated by the above formula, or approximated by the multiple of the difference of two Gaussian windows separated in the x direction:

$$f \approx c(G(x - d/2, \sigma) - G(x + d/2, \sigma))$$

The smaller the displacement d is, the higher multiplication must be used and the more accurate result will be obtained. In the following we investigate what happens if we use as large a displacement as it does not distort the resulting function's shape too much. The purpose in doing so is that the above integral of any feature vector interpreted on points of the image could be integrated by using displaced Gaussian kernels of variance σ_0, followed by taking their linear combinations. This can be used to to evaluate integrals of kernels $G((x - d), \sigma)$ and also $(x - d)G((x - d), \sigma)$. Moreover, utilizing that

$$\partial G(x, \sigma)/\partial\sigma = 1/\sigma((x/\sigma)^2 - 1)G(x, \sigma)$$

functions like $(x - d)^2 G((x - d), \sigma)$ can also be evaluated. The motivation to evaluate weighting functions in this way is that hopefully, a wide range of shifts (d) and sizes (σ) can be quickly approximated using the linear combinations of a few equally spaced integrals.

Extending this to 2D Gaussian functions, components of SAFT can be approximated by linear combinations of $\mathbf{C}$ integrated on different places. Figure 5 shows approximation of SAFT components, and differences from the exact weighting functions. Figure 6 shows arrangement of kernels used to integrate $\mathbf{C}$, which was used to approximate $\mathbf{M}$ of an integration kernel only slightly wider. To approximate $\mathbf{M}$ on a window of σ_{int}, this setup uses 7 Gaussian windows of $\sigma_0 = 0.955 \cdot \sigma_{int}$, separated at a distance of $d_{smpl} = 0.725 \cdot \sigma_{int}$ and arranged in a hexagonal grid. The maximum difference of the weighting functions and their approximation is 4.2%, calculated as a ratio of the quadratic mean (effective) value of the functions and the difference from their approximations. This evaluation enables the approximation of $\mathbf{M}$ on many similar integration windows using only a few equally spaced integrals of $\mathbf{C}$.

5.1 Test Results

This section describes the comparison of standard and alternative evaluation of SAFT. This comparison could be made based on matching or geometric abilities of SAFT. The alternative evaluation is an approximation, which can be considered as adding slight disturbances to the exact SAFT matrix. Matching abilities are robust to small disturbances and performance is signaled by a scalar

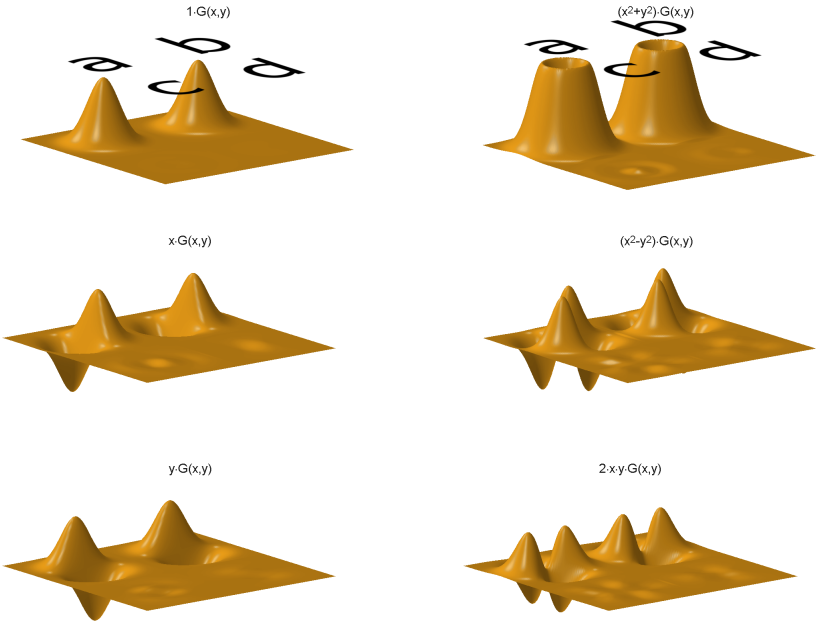

Fig. 5. 6 Weighting functions of SAFT: corresponding to 1, x, y (1^{st} column) and $x^2 + y^2$, $x^2 - y^2$, $2xy$ (2^{nd} column). For each weighting function, accurate(a), approximated(b), difference(c) and accurate compressed to have same effective value as difference(d) is shown.

measure of inlier ratio. Extraction of geometric information is directly affected by disturbances, which are shown by comparing the extracted geometric information superimposed on the images. Therefore, the latter will be used.

During the tests, the image is blurred with $\sigma_{diff} = 1.3pixels$, then image gradients and and $[g_x^2 \;\; g_x g_y \;\; g_y^2]$ (Harris-image) are computed on each pixel. $\mathbf{C}$ is then integrated with basis functions of $G([x \;\; y]^\mathsf{T}, \sigma_{int}) \cdot [x^2 \;\; xy \;\; y^2 \;\; x \;\; y \;\; 1]$ to obtain the 3×6 matrix $\mathbf{M}_e$ which contains the independent elements of exact SAFT matrix $\mathbf{M}$. To obtain the approximation, the Harris-image is blured with σ_0. The result is sampled at 7 positions corresponding to Figure 6, with distance between samples of d_{smpl}. The obtained data is organized into a 3×7 matrix

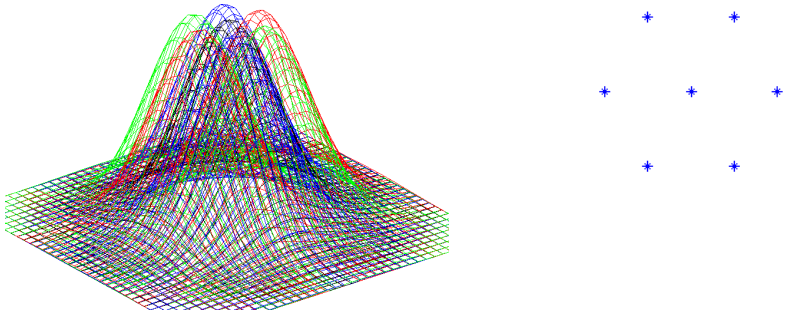

Fig. 6. Hexagonal arrangement of 7 Gaussian window used for the approximate evaluation of SAFT matrix $\mathbf{M}$.

$\mathbf{M}_a$, which is transformed by

$$\mathbf{M}'_a = \mathbf{M}_a\mathbf{T}, \mathbf{T} = \begin{bmatrix} 1.9326 & 0.0000 & -0.6123 & -0.4810 & -0.0000 & 0.0581 \\ 0.0301 & -1.1034 & 1.2977 & -0.2427 & 0.4180 & 0.0585 \\ 0.0301 & 1.1034 & 1.2977 & -0.2427 & -0.4180 & 0.0585 \\ -2.9944 & -0.0000 & -2.9744 & 0.0000 & -0.0000 & 0.6516 \\ 0.0301 & 1.1034 & 1.2977 & 0.2427 & 0.4180 & 0.0585 \\ 0.0301 & -1.1034 & 1.2977 & 0.2427 & -0.4180 & 0.0585 \\ 1.9326 & -0.0000 & -0.6123 & 0.4810 & 0.0000 & 0.0581 \end{bmatrix}.$$

to obtain independents elements of the approximated SAFT descriptor. $\mathbf{T}$ was determined by minimizing squared differences between approximated and original basis functions, see Figure 5. Then $\mathbf{M}$ is reconstructed from both $\mathbf{M}_e$ and $\mathbf{M}'_a$ and several geometric information is extracted.

Figure 7 shows these geometric items superimposed on the original images. Considerable difference is seen only on subfigures 6,7 and 8 regarding the detected parabolae (blue). Inaccuracy of the exact method is due to the small size of the Gaussian window used. The approximation performs worse in case 6, but better for image 8. The approximation is not distinguishable from the exact method in other cases shown. It can concluded that the alternative method using 7 samples can be used for practical applications. Applications which require accurate geometric information extraction should use more samples, but further investigation has to be made regarding accuracy.

5.2 Possible applications

Several modifications of algorithms related to SAFT can be based on this results. For example, the Harris image can be blurred (with σ_0) and sampled over a

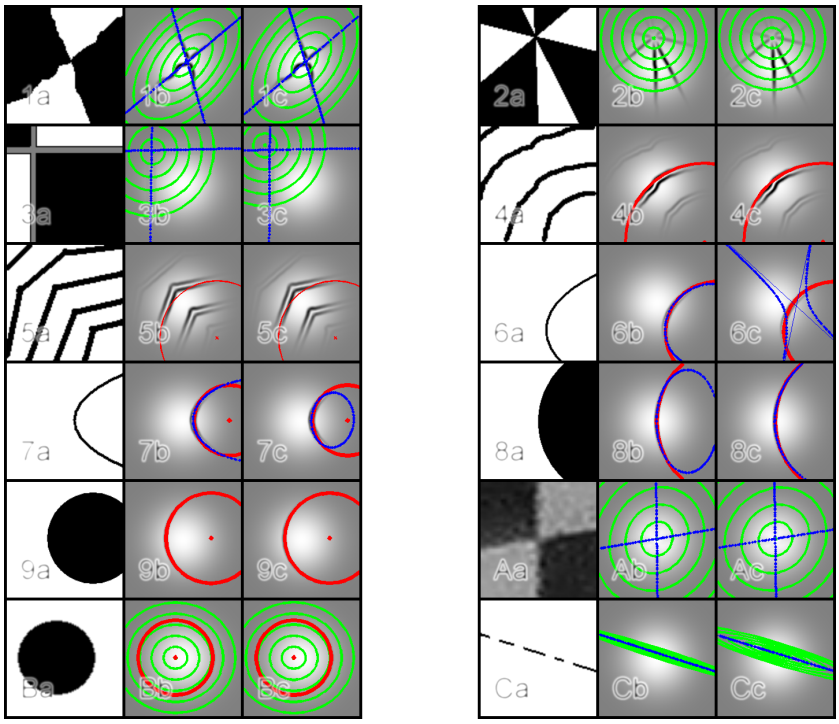

Fig. 7. Comparison of extracted geometric quantities. (a): Original Image (b,c):Gradient strength image windowed with applied Gaussian filter (of $\sigma = 0.2\ imagewidth$). Superimposed geometric quantities are calculated by SAFT matrix extracted by definition (b) and by approximation (c). Green: Potential lines of accumulated velocity distribution ($\mathbf{W}_{\mathsf{H}}$). Red: Center of best rotation and mean radius. Blue: Approximated conic for shapes of rank($\mathbf{M}$) = 4.

hexagonal grid of lattice size $d_{smpl}/2$. Then, the SAFT matrix can be evaluated at any position with 21 read operations and with a 7×6 matrix-multiplication. Applications which require calculation of the SAFT matrix over overlapping windows of the same size include vectorization of pixelized images and search for SAFT-characteristic features or regions fixed against affine distortions (which are characterized by (19), (20) and (21)). When the whole image plane has to be analyzed, the blurred Harris-image can be obtained by a method similar to building of Gaussian Scale Spaces. Even the evaluation of the SAFT matrix of the same differentiating kernel but larger integrating kernels can be speed up by using blurred Harris-images. It is useful only if many SAFT matrix has to be extracted, so the fastening of the calculation of individual SAFT descriptors is larger than the computational demand of calculating the blurred Harris-images. However, these calculations fit to the abilities of GPUs, which perform uniform calculations over uniform data-sets much more efficiently than CPUs. Extraction

of thousands of SAFT descriptors for photograph matching and affine-shape adaptation are expected to utilize this possibility.

6 Conclusions

The connection of the Harris detector-based Affine Adaptation (Harris-Affine) and the Self Affine Feature Transform (SAFT) was analyzed in this paper. It was shown, that due to the derivative rules of planar Gaussian functions, the information enclosed to the SAFT descriptor describes the linear dependence of the Harris matrix (or second moment of gradients) from the 5 free parameters of the used Gaussian integration kernel. It was shown that this relation can be used to decrease the computational requirement of the Harris-Affine iteration, however, further research has to be made in this direction. Based on the connections shown, an alternative method was introduced which evaluates the SAFT matrix. Application which can utilize this alternative descriptor extraction were enumerated. As aconsequence, properties of the SAFT matrix of regular shapes were connected to affine-adaptation.

Acknowledgements

The research in this article was supported by the Hungarian National Research Program grant No. OTKA K 71762 and. This work is connected to the scientific program of the " Development of quality-oriented and harmonized R+D+I strategy and functional model at BME" project, (Project ID: TAMOP-4.2.1/B-09/1/KMR-2010-0002).

References

1. Prohaszka, Z.: Affine invariant features from self-flow. In: RAAD 17th International Workshop on Robotics in Alpe-Adria-Danube Region. (2008) 1–10
2. Prohaszka, Z.: Matching image details with the self affine feature transform. In: V. Magyar Szamitogepes Grafika es Geometria Konferencia. (2010) 206–213
3. Prohaszka, Z., Lantos, B.: Extracting geometric information from images with the novel self affine feature transform. Accepted to: Periodica Polytechnica, Electrical Engineering (2010)
4. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60** (2004) 91–110
5. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence **10** (2005) 1615–1630
6. Tuytelaars, T., Van Gool, L.: Wide baseline stereo matching based on local, affinely invariant regions. In: Proc. of BMVC-00. (2000) 412–425
7. Lindeberg, T.: Detecting salient blob-like image structures and their scales with a scale-space primal sketch:a method for focus-of-attention. International Journal of Computer Vision **11** (1993) 283–318
8. Lindeberg, T.: Scale-space theory: A basic tool for analysing structures at different scales. Journal of Applied Statistics **21** (1994) 225–270

9. Lindeberg, T., Garding, J.: Shape-adapted smoothing in estimation of 3-d shape cues from affine deformations of local 2-d brightness structure. Image and Vision Computing **15** (1997) 415–434
10. Mikolajczyk, K., Schmid, C.: An affine invariant interest point detector. In: Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark, Springer (2002) 128–142 Copenhagen.
11. 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 **65** (2005) 43–72
12. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide-baseline stereo from maximally stable extremal regions. In: Proc. of BMVC-02. (2002) 384–393
13. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Imaging Understanding Workshop. (1981) 121–130