Maryna Mukhina

COMPARISON OF ERROR METRICS IN MATCHING ALGORITHMS
OF IMAGES BY SURF DETECTOR

National Aviation University, Kyiv, Ukraine
1, Kosmonavta Komarova avenue, Kyiv, 03680, Ukraine
E-mail: m_mukhina@inbox.ru

Abstract. Speed-Up Robust Feature (SURF) method is used to detect feature points of images. The analyses of
matching algorithms of feature points is done. The comparison of different error metrics by their accuracy and
computing efficiency is provided on the series of test images for basic transforms like scaling, rotation and shifting.

Keywords: error metric; feature point; homography matrix; normalized cross-correlation; Speed-Up Robust Feature.

Introduction

Matching of images is used in variety of practical
applications like photomapping by unmanned aerial
vehicles (UAV), visual correlation-extreme
navigation, target detection, etc. Most of these
application require robust real-time algorithms to
detect the feature points and then to provide their
reliable matching. In some situations, e.g., for video
sequences or for stereo pairs that have been rectified
the local motion around each feature point may be
mostly translational. In this case, simple error
metrics, such as the sum of squared differences or
normalized cross-correlation can be used to directly
compare the intensities in small patches around each
feature point. Because feature points may not be
exactly located, a more accurate matching degree
can be computed by performing incremental motion
refinement but this can be time consuming and can
sometimes even decrease performance.

Among method of feature detection the scale
invariant feature transform (SIFT) [3] is at present a
very popular basis for image stitching. SIFT delivers
point-wise correspondences between distinctive,
non-repetitive local features in the two images. The
number of detected features is significantly smaller
than the number of pixels in the image. Other
methods for identifying features include local image
descriptors like intensity patterns [5] and the
Kanade-Lucas-Tomasi Feature Tracker (KLT) [6].

But for real-time application Speed-Up Robust
Feature (SURF) method is used [1], since it
approximates or even outperforms previously
proposed schemes with respect to repeatability,
distinctiveness, and robustness, yet can be computed
and compared much faster.

2. Problem statement

Much of the performance increase in SURF can be
attributed to the use of an intermediate image
representation known as Integral Image. The integral
image is computed rapidly from an input image and
is used to speed up the calculation of any upright
rectangular area. Given an input image \( I \) and a point
\((x,y)\) the integral image is calculated by the sum of
the values between the point and the origin.

Formally this can be defined by the formula:

\[
I_x(x,y) = \sum_{i=0}^{x} \sum_{j=0}^{y} I(i,j)
\]  

(1)

Using the integral image, the task of calculating
the area of an upright rectangular region is reduced
to four operations.

Since computation time is invariant to change in
size this approach is particularly useful when large
areas are required. SURF makes good use of this
property to perform fast convolutions of varying size
box filters at near constant time.

The SURF detector is based on the determinant
of the Hessian matrix:

\[
H(x,y,\sigma) = \begin{bmatrix}
L_{xx}(x,y,\sigma) & L_{xy}(x,y,\sigma) \\
L_{xy}(x,y,\sigma) & L_{yy}(x,y,\sigma)
\end{bmatrix}.
\]

(2)

where \( L_{xx}(x,y,\sigma) \) – convolution of second order partial
of Gaussian \( \frac{\partial^2}{\partial x^2} g(\sigma) \) from the image in the point \((x,y)\).
The same is for \( L_{xy}(x,y,\sigma) \) and \( L_{yy}(x,y,\sigma) \).

But for SURF the fast Hessian is found that is the
approximation of matrix (2) by box filters. Dimension of filters is selected as 9x9 with scale
\( \sigma = 1.2 \) (minimal). The approximations are
designated as \( D_{xx}, D_{yy}, D_{xy} \). The weights are
selected from Frobenius norm:

\[
\det(H_{approx}) = D_{xx}D_{yy} - (0.9D_{xy})^2.
\]
In general case the descriptor of feature point by SURF method includes the following information: coordinates \( P = \{x, y\} \), scale of Gaussian filter \( M = \{\sigma\} \), gradient orientation \( R = \{\theta\} \), Laplacian \( L = \{0,1\} \) (means either white spot on black background or black spot on white), and gradients of quadrants \( D = \{D_1, D_2, ..., D_{64(128)}\} \), which surround the point.

To calculate the descriptor the rectangular area is formed around the feature point. It has the size \( 20\sigma \), where \( \sigma \) – filter scale, that was used to find the point. For the first octave the size of area is 40x40 pixels. The quadrant is oriented along the major direction calculated for feature point.

The descriptor is calculated as the gradients for \( 4 \times 4 = 16 \) quadrants around the feature point. Then each quadrant is divided further by 16 smaller quadrants as it is shown in Fig. 1.

![Fig. 1. Descriptor of feature point [1]](image)

For each quadrant the responses of Haar wavelets of size \( 2\sigma \) are computed on the regular grid \( 5 \times 5 = 25 \). Responses by directions \( x \) and \( y \) are designated as \( dx \) and \( dy \), respectively, and then for each quadrant the following vector is found:

\[
D_{\text{quadrant}} = \left[ \sum dx, \sum dy, \sum |dx|, \sum |dy| \right].
\]

With Haar wavelets calculation the image is not rotated, the filter is computed in image coordinates. But after the gradients coordinates \( dx, dy \) are rotated in angle corresponding to orientation of quadrant.

Four components on each quadrant must be computed that gives totally the 64 components of descriptor of area around the feature point. By the forming of descriptor array the values are weighted by Gaussian \( 3.3\sigma \) and centered in the feature point to minimize the possible noise components.

After detecting all feature points on the pair of compared images it is necessary to find matches between these points.

3. Related works

Matching of feature points found by any descriptors is performed usually by well known method as random sampling (RANSAC) researched e.g. for SURF algorithm in [2]. Transformation between the pair of images can be described by homography matrix \( 3 \times 3 \):

\[
\begin{bmatrix}
    a_{11} & a_{12} & a_{13} \\
    a_{21} & a_{22} & a_{23} \\
    a_{31} & a_{32} & a_{33}
\end{bmatrix},
\]

which describes the transformation of image point \((x, y)\) into the point \((x', y')\) on the other image by the following relationships:

\[
\begin{align*}
x' &= a_{11}x + a_{12}y + a_{13} \\
y' &= a_{31}x + a_{32}y + a_{33}.
\end{align*}
\]

The calculation of image renovation needs to work out at least eight parameters, and at least eight equations in theory are needed, so it needs at least 4 non-collinear feature points. False matching points will influence the result of the least squares estimates.

RANSAC algorithm steps:

1) From the equation of matching feature points are selected, randomly selected from 4 points to set up equations, solve the eight unknown parameters of matrix \( H \).

2) Calculate rest of the feature points after matrix \( H \) transformation, and calculate the distance between candidate matching points.

3) If the distance is less than a certain value, the candidate point is looked as interior point or the outside point.

4) Make the statistics of the quantity of interior point under the homography matrix.

5) Choose another four match points, carry out steps 1 to 4 again, and repeat several times, choose the collection with largest number of interior points.

Other approach is based on the fast approximate nearest neighbors (FLANN) algorithm investigated in [4]. It uses the nearest distance as Euclidean distance, defined as following:

\[
D = \sqrt{(x_1 - x'_1)^2 + (x_2 - x'_2)^2 + ... + (x_{64} - x'_{64})^2},
\]

where \((x_1, x_2, ..., x_{64})\), \((x'_1, x'_2, ..., x'_{64})\) are detected features of points to be matched.
Any of these methods require the error metrics to be used in approximation or sampling. Calculation of error metrics must be fast and stable, with good repeatability. Geometrical correctness refers to the point being localized on the same image structure or region and can be measured in a number of ways. A simple approach is to evaluate the correspondences by eye, but this is laborious and lacks robustness. Another method is used that relates the points by a homography. Homography matrix maps points \( X \) in one image to the points \( X' \) in another. Using this function it is possible to determine whether a point has a correspondence by performing the mapping operation and checking for detected points within a neighborhood of the target location.

### 4. Error metrics comparison

Three main error metrics are selected to compare their efficiency by SURF method: sum of absolute differences (SAD)

\[
E_{SAD} = \sum |D_i - D_j|,
\]

sum of squared differences (SSD)

\[
E_{SSD} = \sum (D_i - D_j)^T (D_i - D_j),
\]

and normalized cross-correlation (NCC)

\[
E_{NCC} = \sum D_i^T \cdot D_j,
\]

where \( D_i, D_j \) - matrices of detected feature points \( i \) and \( j \) on the pair of images. It is obvious that in all of the cases the error metric will be the matrix of dimension \( n \)-by-\( m \), where \( n \) and \( m \) - are numbers of feature points detected on both compared images.

Realization of SURF method in (Code of SURF listing in MATLAB) was used in practice for experiments. The descriptors are formed as matrix \( D \) by size \( 64 \times N \), or \( 128 \times N \), where \( N \) – number of feature points. The function (7) can be found by single multiplication of two matrixes of compared image descriptors:

\[
E_{NCC} = D_i^T \cdot D_j = \begin{bmatrix}
D_1^T & \cdots & D_N^T
\end{bmatrix}\begin{bmatrix}
D_1 \\
\vdots \\
D_N
\end{bmatrix},
\]

where

\[
E_{SAD} = \begin{bmatrix}
D_1 - D_1 \\
\vdots \\
D_N - D_N
\end{bmatrix},
\]

\[
E_{SSD} = \begin{bmatrix}
(D_1 - D_1)^T (D_1 - D_1) \\
\vdots \\
(D_N - D_N)^T (D_N - D_N)
\end{bmatrix},
\]

Since descriptor matrix is already normalized due to peculiarities of SURF method, then it is possible to state that each component of matrix \( E_{NCC} \) is normalized. Search of maximal elements is done by the finding the maximal values of matrix (8) in each row and checking whether this value is greater than the threshold. Only one maximal element in each row is selected since it is supposed that one point on the template will be matched to the single point on the current image.

Error metrics (5), (6) are realized as vector difference and multiplication in loops by indexes \( n \) and \( m \) in order to form the same matrices \( E_{SAD} \) and \( E_{SSD} \).

### 5. Experimental results of error metrics comparison

Set of images with known homography matrices was used [7]. Testing of correctness of detected feature points on the pair of images was done by all type of error metrics (5), (6), (7). Tests were done using known homography matrices of images. Results are represented in Tables 1, 2, 3 as the ratios between the general number of detected features \( N \), number of true matched points \( N_{true} \), number of false matched points \( N_{false} \), time of calculation of each type of error metrics. Tests were done in MATLAB 7.8.0. Examples of matching are shown in Fig. 2, Fig. 3.

**Fig. 2.** Matching of two images (a-2) with homography matrix \( H = [8.5828552e-01, 2.1564369e-01, 9.9101418e+00;-2.1158440e-01, 8.5876360e-01, 1.3047838e+02; 2.0702435e-06, 1.2886110e-06, 1.0000000e+00] \)

**Fig. 3.** Matching of two images (b-2) with homography matrix \( H = [8.5828552e-01, 2.1564369e-01, 9.9101418e+00;-2.1158440e-01, 8.5876360e-01, 1.3047838e+02; 2.0702435e-06, 1.2886110e-06, 1.0000000e+00] \)
Table 1. Results for NCC metrics (threshold value is 0.97)

<table>
<thead>
<tr>
<th>Pair of images</th>
<th>$N_{true} / N$</th>
<th>$N_{false} / N$</th>
<th>Time, sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>a-2</td>
<td>46/46</td>
<td>0/46</td>
<td>0.2454</td>
</tr>
<tr>
<td>a-3</td>
<td>35/139</td>
<td>104/139</td>
<td>0.2081</td>
</tr>
<tr>
<td>a-4</td>
<td>3/28</td>
<td>25/28</td>
<td>0.1009</td>
</tr>
<tr>
<td>a-5</td>
<td>223/245</td>
<td>22/245</td>
<td>0.3702</td>
</tr>
<tr>
<td>a-6</td>
<td>20/61</td>
<td>41/61</td>
<td>0.1037</td>
</tr>
<tr>
<td>b-2</td>
<td>125/138</td>
<td>13/138</td>
<td>0.2807</td>
</tr>
<tr>
<td>b-3</td>
<td>59/96</td>
<td>37/96</td>
<td>0.2230</td>
</tr>
<tr>
<td>b-4</td>
<td>12/33</td>
<td>21/33</td>
<td>0.1554</td>
</tr>
<tr>
<td>b-5</td>
<td>3/5</td>
<td>2/5</td>
<td>0.1309</td>
</tr>
<tr>
<td>b-6</td>
<td>1/17</td>
<td>16/17</td>
<td>0.1287</td>
</tr>
</tbody>
</table>

Table 2. Results for SSD metrics (threshold value is 0.03)

<table>
<thead>
<tr>
<th>Pair of images</th>
<th>$N_{true} / N$</th>
<th>$N_{false} / N$</th>
<th>Time, sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>a-2</td>
<td>74/74</td>
<td>0/74</td>
<td>4.1528</td>
</tr>
<tr>
<td>a-3</td>
<td>30/100</td>
<td>70/100</td>
<td>3.6245</td>
</tr>
<tr>
<td>a-4</td>
<td>3/28</td>
<td>25/28</td>
<td>2.0733</td>
</tr>
<tr>
<td>a-5</td>
<td>223/245</td>
<td>22/245</td>
<td>2.1784</td>
</tr>
<tr>
<td>a-6</td>
<td>20/61</td>
<td>41/61</td>
<td>1.6149</td>
</tr>
<tr>
<td>b-2</td>
<td>92/102</td>
<td>10/102</td>
<td>2.9622</td>
</tr>
<tr>
<td>b-3</td>
<td>39/67</td>
<td>28/67</td>
<td>3.0226</td>
</tr>
<tr>
<td>b-4</td>
<td>4/15</td>
<td>11/15</td>
<td>3.0305</td>
</tr>
<tr>
<td>b-5</td>
<td>3/5</td>
<td>2/5</td>
<td>2.4859</td>
</tr>
<tr>
<td>b-6</td>
<td>1/17</td>
<td>16/17</td>
<td>2.6992</td>
</tr>
</tbody>
</table>

Table 3. Results for SAD metrics (threshold value is 0.03)

<table>
<thead>
<tr>
<th>Pair of images</th>
<th>$N_{true} / N$</th>
<th>$N_{false} / N$</th>
<th>Time, sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>a-2</td>
<td>68/68</td>
<td>0/68</td>
<td>3.6381</td>
</tr>
<tr>
<td>a-3</td>
<td>20/105</td>
<td>85/105</td>
<td>2.9929</td>
</tr>
<tr>
<td>a-4</td>
<td>4/26</td>
<td>22/26</td>
<td>1.7089</td>
</tr>
<tr>
<td>a-5</td>
<td>230/252</td>
<td>22/252</td>
<td>1.8438</td>
</tr>
<tr>
<td>a-6</td>
<td>109/537</td>
<td>428/537</td>
<td>1.4983</td>
</tr>
<tr>
<td>b-2</td>
<td>122/136</td>
<td>14/136</td>
<td>2.5629</td>
</tr>
<tr>
<td>b-3</td>
<td>54/97</td>
<td>43/97</td>
<td>2.5954</td>
</tr>
<tr>
<td>b-4</td>
<td>10/25</td>
<td>15/25</td>
<td>2.5311</td>
</tr>
<tr>
<td>b-5</td>
<td>4/6</td>
<td>2/6</td>
<td>2.2979</td>
</tr>
<tr>
<td>b-6</td>
<td>0/7</td>
<td>7/7</td>
<td>2.1109</td>
</tr>
</tbody>
</table>

As can be seen from Fig. 4 the accuracy of matching is approximately the same for all three error metrics but time of calculation is significantly smaller for NCC error metrics.

Fig. 4. Comparison of error metrics for tested pairs of images

6. Conclusions

NCC error metrics demonstrates the same results as widely used SSD with significantly being more effective in computing. The time efficiency is about 14 times greater than for both SSD and SAD error metrics that is easily explained by decreasing the number of calculation and excluding the loop iterative operations of vector difference and multiplication and replacing them by single matrix multiplication.

References


Received 10 June 2014
М. П. Мухіна. Порівняння метрик похибок в алгоритмі співставлення зображень детектором surf
Національний авіаційний університет, Київ, Україна, проспект Космонавта Комарова, 1
E-mail: m_mukhina@inbox.ru
Метод Speed-Up Robust Feature (SURF) використовується для робастного виявлення характерних точок зображень. Проаналізовано алгоритми співставлення характерних точок. Проведено порівняння різних метрик похибок за їх точністю та обчислювальною ефективністю на серії тестових зображень для базових перетворень, таких як масштабування, поворот і зсув.
Ключові слова: матриця гомографії; метрика похибки; нормалізована взаємна кореляція; прискорений робастний детектор характерний ознак; характерна точка.

M. P. Mukhina. Сравнение метрик погрешностей в алгоритмах сопоставления изображений детектором surf
National Aviation University, Kyiv, Ukraine, prosp. Kozmonavta Komarova, 1
E-mail: m_mukhina@inbox.ru
Метод Speed-Up Robust Feature (SURF) используется для робастного обнаружения характерных точек изображений. Проанализированы алгоритмы сопоставления характерных точек. Проведено сравнение различных метрик погрешностей по их точности и вычислительной эффективности на серии тестовых изображений для базовых преобразований, таких как масштабирование, поворот и смещение.
Ключевые слова: метрика ошибки, характерная точка, матрица гомографии, нормализованная взаимная корреляция, ускоренный робастный детектор характерных признаков.

Mukhina Maryna
Candidate of Engineering. Associate Professor.
Department of Aviation Computer-Integrated Complexes, National Aviation University, Kyiv, Ukraine.
Education: National Aviation University, Kyiv, Ukraine.
Research area: image processing, correlation-extremal navigation, data fusion.
Publications: 35.
E-mail: m_mukhina@inbox.ru