

UNIFIED TEMPLATE OF GEOREFFERENCED IMAGES FOR VISUAL CORRELATION-EXTREME NAVIGATION SYSTEM

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

Abstract. *The article considers existing methods of image feature detection. Descriptive representation of unified template is proposed based on Speed-Up Robust Feature (SURF) method. The requirements and structure of database of unified template are developed including the reliability factor of each image feature. Experiments evidences the decreasing of memory volume for database and of computational time of recognition.*

Keywords: correlation-extreme navigation; feature point; georeferencing; image recognition; reliability (weight) factor; Speed-Up Robust Feature.

1. Introduction

Correlation-extreme navigation system (CENS) is topical since it is an alternative of data fusion source of Inertial Navigation System (INS), instead of Satellite Navigation System (SNS). SNS allows determining the coordinates of unmanned aerial vehicle (UAV) quite precisely and reliably, but possesses a number of essential shortcomings. The main is that satellite signal can be easily jammed by occasion or intentionally. And then INS operation becomes autonomous with accumulation of positioning errors. The INS correction is required and the most favorable source is CENS. Application of CENS allows carrying out the correction of heading, coordinates and velocity of UAV in the conditions of presence of cartographic data.

The principle of operation of CENS is based on a comparison of template realization of geophysical field (saved in computer memory) with the current realization obtained from the airborne sensor [1]. A degree of matching is represented in the form of correlation function, its extremum is found and is used to get positioning over map.

Visual correlation-extreme navigation consists of two main parts. The first, absolute navigation, or georeferencing [2] mostly relies on recognition of current image frame among template satellite imaginary by feature and correlation approaches. The second, relative navigation, or visual odometry [3] takes a series of video frames and obtains the relative velocities from increments between them. Here also possible two approaches: feature tracking from frame to frame or correlation between consequence frames. Both tasks of visual navigation, recognition and tracking, must be solved

in real-time by processing large amount of video information.

Accuracy of CENS positioning is determined by accuracy of airborne sensor and, of course, by accuracy of cartographic data. The high resolution of georeferencing images for visual CENS provides redundancy of information, requires large volume of memory to be allocated, needs high computational capability to be processed in real time, etc. Thus, the preliminary processing of georeferencing images is recommended with purpose of creation of unified template. It must contain minimal but sufficient data for reliable matching with heavy computational preprocessing.

2. Problem statement

The georeferencing image is represented by a descriptive set containing image features like points, contours, areas, textures. The feature points detected by SURF method [4] are selected for investigation.

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

$$\mathbf{H}(x, y, \sigma) = \begin{vmatrix} L_{xx}(x, y, \sigma) & L_{yx}(x, y, \sigma) \\ L_{xy}(x, y, \sigma) & L_{yy}(x, y, \sigma) \end{vmatrix}, \quad (1)$$

where $L_{xx}(x, y, \sigma)$ is convolution of the 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 (1) by box filters. Dimension of filters is selected as 9×9 with scale $\sigma = 1.2$ (minimal). The approximations are designated

as D_{xx} , D_{yy} , D_{xy} . The weights are selected from Frobenius norm:

$$\det(\mathbf{H}_{approx}) = D_{xx}D_{yy} - (0.9D_{xy})^2.$$

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 $R = \{\alpha\}$, Laplacian $L = \{0, 1\}$ (means either white spot on black background or black spot on white), and gradients of quadrants $D = \{D_1, D_2, \dots, D_{64(128)}\}$, which surround the point.

To calculate the descriptor the rectangular area is formed around the feature point. It has the size 20σ , where σ - filter scale, that was used to find the point. For the first octave the size of area is 40×40 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.

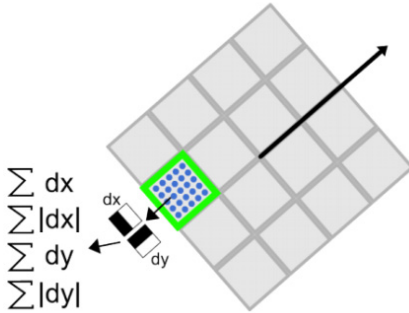


Fig. 1. Descriptor of feature point [4]

For each quadrant the responses of Haar wavelets of size 2σ 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:

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

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 3.3σ and centered in the feature point to minimize the possible noise components.

The descriptor of feature point by SURF contains enough information to match the reference image and current one, but for reliable positioning some fields to the descriptor must be added and investigated.

3. Related works

General requirements to unified template for CENS are formulated in [5], without any proposed structure and used methods of data processing. More detailed description of procedure of template data preparation and processing is given in [6]. The proposed presentation is based on technology of computer vision where the template is given as a scene. However, such representation has a significant drawback, namely the procedure of its formation cannot be fully formalized. And therefore it has no possibility to be fully automatized.

4. Proposed database structure of unified template

Meta template (georeferenced image) is separated into several fragments (number and sizes are needed to be researched and optimized also), or templates with the unique ID to be easily matched over meta image, and then they are processed by SURF detector. Coordinates of feature points detected by SURF are given in image coordinate system (CS) and must be referenced to geographical CS, which requires two additional fields for latitude φ and longitude λ . As an alternative, it is possible to set the coordinates of reference point (image center of left upper point) with corresponding coordinate transformation matrix (DCM - direction cosine matrix). The last variant provides minimization of memory volume required to keep unified template data. It requires coordinates (φ , λ) of reference point in geographical CS, size of template either in meters of pixels ($n \times m$), and DCM (\mathbf{B}).

Also, two additional navigation parameters must be incorporated into unified template for visual CENS: heading ψ (that is, orientation of image) and altitude H (that is, image scale). They are the same for all feature points and therefore together with geographical coordinates of reference image point and with coordinate transformation matrix can be positioned to separate data level of template (Fig. 2).

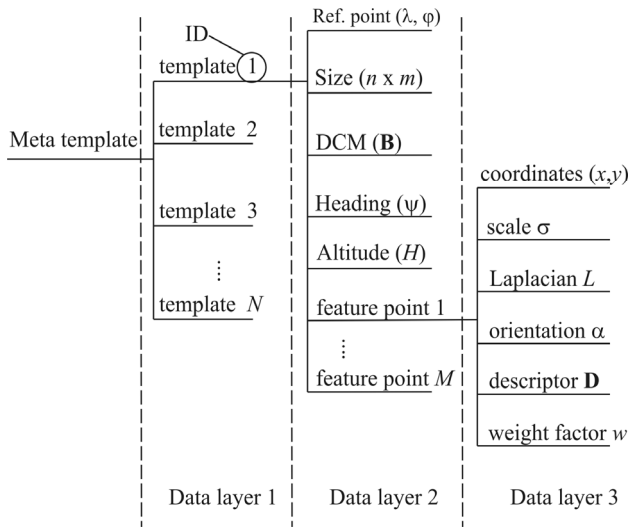


Fig. 2. Database structure of unified template

The next data layer contains information about each feature point, the same as for conventional SURF descriptor, but including one more, important field, namely weight factor.

SURF detector provides redundant number of feature points [4], the most of which have a lack of repeatability under image distortions or noises. Obviously, it is necessary to test the points during template preprocessing and select the most reliable, that is, those to be detected and matched over all distortions and noises.

The intuitive determination of weight factor of i^{th} feature point can be represented as

$$w_i = \frac{N_i}{N}, \tag{2}$$

where N_i is the number of reliable matching of feature point over all N distortions and noises of a template.

The error metric of feature point matching is selected as normalized cross-correlation coefficient (NCC) with maximal value to be equal to 1 if descriptors of two points are equivalent, and to zero if they are totally different. NCC is easy to calculate for SURF descriptor [7] since it requires single operation of matrix multiplication.

5. Experimental results of error metrics comparison

A set of images obtained from UAV camera is investigated. An image of 4000x3000 px is divided into N templates (number of template is selected as 4). Each template correspondingly has the resolution 2000x1500 px.

SURF detector is realized in MATLAB 2014a environment by [8]. For the first template of meta georeferenced image shown in Fig. 2, a, the initial data fields of feature point descriptors are represented in Fig. 2, b.



a

Fields	x	y	scale	laplacian	orientation	descriptor
1	295.8586	12.4421	1.9188	1	4.3786	64x1 double
2	391.5719	12.3750	1.9518	1	4.2510	64x1 double
3	457.6401	12.9136	2.2514	0	5.1500	64x1 double
4	466.2005	11.7386	2.1497	1	5.7199	64x1 double
5	602.3679	12.7518	2.0557	1	0.5578	64x1 double
6	628.1555	12.1941	1.7805	1	5.3728	64x1 double
7	1.0285e+03	11.8624	2.1594	1	2.2321	64x1 double
8	1.4277e+03	11.3028	2.3028	1	2.6547	64x1 double
9	1.4440e+03	12.4391	1.9043	0	5.0879	64x1 double
10	1.7796e+03	12.2357	1.6026	1	4.6517	64x1 double
11	505.9628	14.9987	2.0396	1	3.2641	64x1 double
12	571.6945	13.9799	1.9515	0	1.3630	64x1 double
13	594.3373	13.9322	2.0245	0	2.7869	64x1 double
14	741.5762	14.1280	1.9871	1	4.2162	64x1 double

b




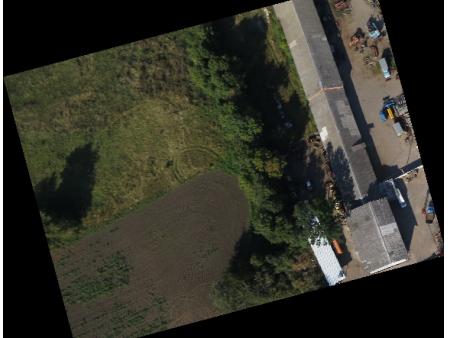

Fig. 2. The meta template and data fields of SURF descriptor of first template points

Also the number of detected feature points for the first template is significant, 9052. And some of this points are not reliable meaning that for other condition of image capturing or illumination they will not be detected at all. Moreover, non-stationary objects like cars, shadows, etc. also need to be

eliminated from a template. For this purpose it is proposed to introduce the weight factor of each point.

For this, the selected template (No.1) was subjected to different distortions and noises (Table 1).

Table 1. Noises and distortions of template image

Darkening with intensity 0.1 MATLAB : <code>imnoise(I01,'gaussian', -0.1,0)</code>		
Darkening with intensity 0.2 MATLAB : <code>imnoise(I01,'gaussian', -0.2,0)</code>		
Darkening with intensity 0.3 MATLAB : <code>imnoise(I01,'gaussian', -0.3,0)</code>		
Lighting with intensity 0.1 MATLAB : <code>imnoise(I01,'gaussian', 0.1,0)</code>		
Lighting with intensity 0.2 MATLAB : <code>imnoise(I01,'gaussian', 0.2,0)</code>		
Lighting with intensity 0.3 MATLAB : <code>imnoise(I01,'gaussian', 0.3,0)</code>		
High frequency gaussian noise with variance 0.05 MATLAB : <code>imnoise(I01,'gaussian', 0,0.05)</code>		
High frequency gaussian noise with variance 0.1 MATLAB : <code>imnoise(I01,'gaussian', 0,0.1)</code>		
Image rotation by 5 degrees MATLAB : <code>imrotate(I01, 5)</code>		
Image rotation by 10degrees MATLAB : <code>imrotate(I01, 10)</code>		
Image rotation by 15 degrees MATLAB : <code>imrotate(I01, 55)</code>		
Affine transforms, vertical skew MATLAB affine transformation matrix <code>[1 0 0; .5 1 0; 0 0 1]</code>		
Affine transforms, horizontal skew <code>[1 0.5 0; 0 1 0; 0 0 1]</code>		
Affine transforms, horizontal stretching <code>[2 1 0; 0 1 0; 0 0 1]</code>		
Affine transforms, vertical stretching <code>[1 1 0; 0 2 0; 0 0 1]</code>		

The reliability of each point was calculated by (2) for general number of experiments $N=15$. Initially the threshold of NCC was selected as 0.75. Fig. 3 demonstrates the distribution of reliable points over the first template.

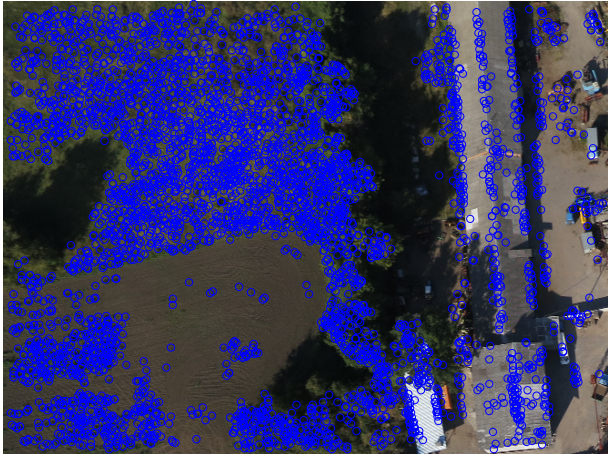


Fig. 3. NCC for reliable matching > 0.75 , number of reliable feature points is 4142 from total number of 9052.

The results of image recognition for other distortions parameters have shown poor results, that is why the threshold of NCC is decided to be increased up to 0.9. Reliable feature points are then shown in Fig. 4.

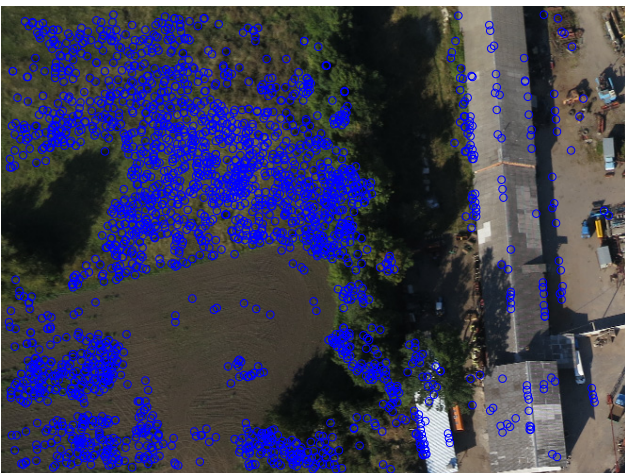


Fig. 4. NCC for reliable matching > 0.9 , number of reliable feature points is 2328 from total number of 9052

The structure of unified template data fields realized in MATLAB is shown in Fig. 5.

Image recognition quality has been estimated by taking the meta image under various distortions like rotation (in the range from $+15$ to 15 deg with 5 deg step), affine transforms (skewing and stretching in horizontal and vertical directions, 4 transforms in

total), illumination changes (darkening and lighting in the range $+0.3$ to -0.3 with 0.1 value step).

High frequency noise has been eliminated from consideration since it is easily filtered at image pre-processing stage. Total number of distortions has been taken correspondingly equal to 16.

For image represented in Fig. 2, a, the total number of SURF points is 43325, causing the descriptor matrix in dimension of 64×43325 . The full search by the meta template using NCC error metric requires matrix multiplication by $43325 \times M$,

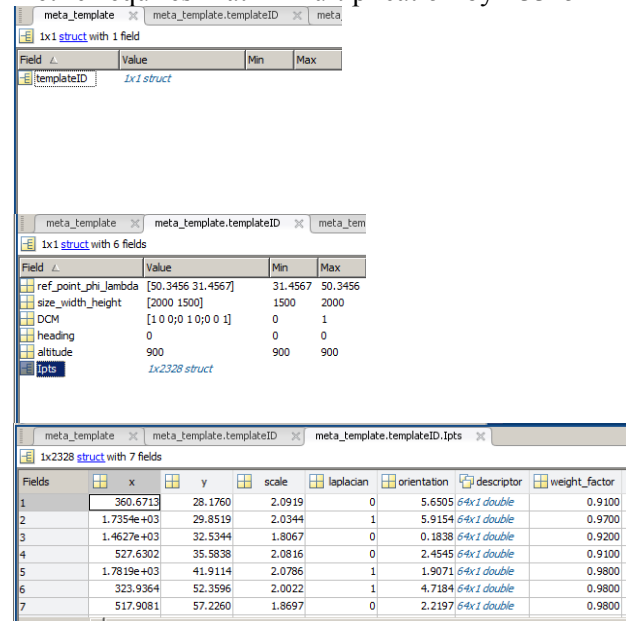


Fig. 5. Proposed modification of structure of template data fields

where M is a number of SURF points found in the captured video fragment. Excluding the time required for SURF detector, the computational cost of the given operation is too high to be used in real time, even for preliminary region recognition. In average, computational time is about 100 sec.

Introducing the proposed unified template allows us to reduce the number of SURF points, in general for meta image it is 6997, meaning decreasing in 6 times. Correspondingly, time for full search over templates is about 2 sec, by several orders of magnitude.

The descriptor matrix D_0 is formed by reshaping descriptors of SUR points (vectors 64×1) into matrix which dimension is $64 \times (N_1 + N_2 + \dots + N_k)$, where N_1, N_2, \dots, N_k are number of SURF points in each k templates selected for search area.

If a prior data is available about current position at meta template, then the dimension of descriptor matrix can be even more reduced.

In experiments the fragment 800px×800px of meta image has been randomly selected (~20% of the size of each template). The results of recognition without distortions have shown poor quality (low

then 60%) since the overlapped area has been not enough. Increasing the captured fragment area up to 50% (size 1200px×1200px) allows improving the quality up to 70% (Fig. 6).

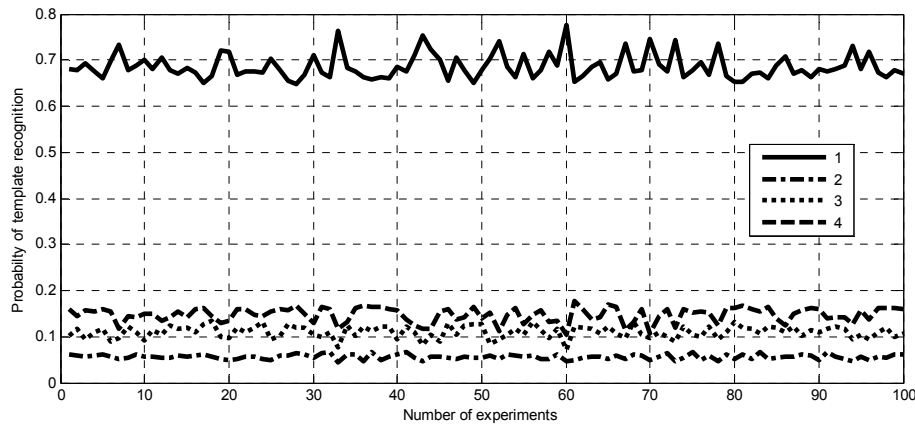


Fig. 6. Recognition quality of video fragment 1200px×1200px randomly taken from the first template

Probability of fragment recognition among the given number of templates has been taken as total number of SURF feature points in unified meta template matched with points detected at fragment. The frequency sense of such probability needs to be normalized to the range [0; 1] then the following expression has been used:

$$h_i = \frac{H_i}{\sum_{i=1}^k H_i},$$

where H_i is number of matched points with i^{th} template, k is total number of templates included in search.

6. Conclusions

Proposed approach to creation of unified template based on SURF points provides minimization of memory volume required to store the cartographic database. Heavy preprocessing stage is used to test and determine the reliability factor of feature points and to select those with reliability higher than predetermined threshold. Decreasing of computational time required for full search over unified templates is about 2 sec, by several orders of magnitude in comparison with search over ordinary template. However, the recognition quality suffers and drops from 90% to 70%.

References

[1] M.P. Mukhina, I.V. Seden Analysis of modern correlation extreme navigation systems //

Electronics and Control Systems, №1 (39), 2014. – p. 95 - 101.

[2] G. Conte and P. Doherty An integrated UAV navigation system based on aerial image matching. In IEEE Aerospace Conference, Big Sky, Montana, March 2008.

[3] Dong-Gyu Sim, Rae-Hong Park, Rin-Chul Kim, Sang Uk Lee, Ihn-Cheol Kim Integrated Position Estimation Using Aerial Image Sequences // IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, No. 1, January 2002. – P. 1-18.

[4] H. Bay, T. Tuytelaars, L. Van Gool. SURF: Speeded up robust features. European Conference on Computer Vision, Vol. 1, pp.404-417, 2006.

[5] Correlation-extreme methods of navigation and of guidance / V.K. Baklitskiy. - Tver': TO "Knizhnyi klub", 2009. - 360 p. (In Russian).

[6] Veremeenko K.K. et al. Modern information technologies in the problems of navigation and guidance of unmanned maneuvering aerial vehicles / Edited by M.N. Krasilshchikov, G.G. Serebriakov. - M.: FISMATLIT. - 2009. - p. 556. (In Russian).

[7] V. Kharchenko, M. Mukhina Correlation-extreme visual navigation of unmanned aircraft systems based on speed-up robust features // Aviation. No 2, volume 18, 2014.

[8] Code of SURF listing in MATLAB. Available from Internet: <http://www.mathworks.com/matlabcentral/fileexchange/28300-opensurf-including-image-warp>

Received 10 September 2015.

М.П. Мухіна. Уніфікований еталон геоприв'язаних зображень для візуальної кореляційно-екстремальної навігаційної системи

Національний авіаційний університет, просп. Космонавта Комарова, 1, Київ, Україна, 03680

E-mail: m_mukhina@inbox.ru

Проаналізовано наявні методи виділення характерних ознак зображення. Запропоновано дескриптивне представлення уніфікованого еталону на основі методу Speed-Up Robust Feature (SURF). Розроблені вимоги і структура бази даних уніфікованого еталону, включаючи показник надійності кожної характерної ознаки зображення. Експерименти свідчать про зменшення обсягів пам'яті для зберігання бази даних і зниження обчислювальних витрат процес розпізнавання.

Ключові слова: геоприв'язка; кореляційно-екстремальна навігація; показник надійності (вага); прискорене виділення робастних характерних ознак; розпізнавання зображень; характерна точка.

М.П. Мухина. Унифицированный эталон геопривязанных изображений для визуальной корреляционно-экстремальной навигационной системы

Национальный авиационный университет, просп. Космонавта Комарова, 1, Киев, Украина, 03680

E-mail: m_mukhina@inbox.ru

Проанализированы существующие методы выделения характерных признаков изображения. Предложено дескриптивное представление унифицированного эталона на основе метода Speed-Up Robust Feature (SURF). Разработаны требования и структура базы данных унифицированного эталона, включая показатель надежности каждого характерного признака изображения. Эксперименты свидетельствуют про уменьшение объемов памяти для хранения базы данных и снижение вычислительных затрат на процесс распознавания.

Ключевые слова: геопривязка; корреляционно-экстремальная навигация; показатель надежности (вес); распознавание изображений; ускоренное выделение робастных характерных признаков; характерная точка.

Maryna Mukhina (1979). PhD, Associate Professor.

Post Doctoral, National Aviation University, Kyiv, Ukraine.

Department of Aviation Computer-Integrated Complexes, National Aviation University, Kyiv, Ukraine

Education: BSc, MSc (1996-2002) in Electrical Engineering, National Aviation University.

Research area: correlation-extreme navigation systems, data fusion algorithms.

Publications: 35.

E-mail: m_mukhina@inbox.ru