TRANSPORT SYSTEMS

UDC 629.051:629.7.014-519 (045)

[1]**M. P. Mukhina,**
[2]**A. P. Prymak,**
[3]**A. N. Babeniuk**

ALGORITHM OF FORMING AND SELECTING OF INFORMATIVE FEATURES IN CORRELATION EXTREME NAVIGATION SYSTEM DATABASE

Aviation Computer-Integrated Complexes Department, National Aviation University, Kyiv, Ukraine
E-mails: [1]mukhina@nau.edu.ua, [2]primak.artem@gmail.com, [3]anitabaly@mail.ru

Abstract*—Hierarchical approach to data fusion in complex navigation system is proposed. Low level of data processing in correlation extreme navigation system includes forming and selecting the most informative relevant features. Advanced algorithm is therefore developed based on weighting correlation sampling of template features. Practical efficiency of algorithm is researched based on Speed-Up Robust Feature method of feature points detection and description. Minimization of feature points number in template without decreasing the quality of matching is observed along with reduction of computational costs.*

Index Terms—Feature selection; maximum entropy method; correlation weighting; Speed-Up Robust Feature; template informativity.

I. INTRODUCTION

As mentioned by ICAO Committee, Strapdown Inertial Navigation System (SINS) is the most informative systems and together with Satellite Navigation System (SNS) allows obtaining all required parameters for object control, including information about angular orientation. SINS is fully autonomous, meaning that it can function without use of any data from other systems. Another advantage of SINS is high sampling rate to output navigation data for the external consumers: sampling rate for orientation angles is up to 100 Hz and for navigation data it is from 10 to 100 Hz. At the same time, SINS cannot be used long time in autonomous mode. Measuring elements of SINS, like gyroscopes and accelerometers have methodical and instrument errors, initial alignment errors are also present and cannot be avoided and together with quantazing errors result in significant accumulation of navigation errors with time. To minimize the errors the standard integration of SINS with external systems is used.

A promising alternative to use of SNS in the integration scheme can be variant of obtaining correction from Correlation Extreme Navigation System (CENS) [1]. Several strong advantages of CENS like quasi independence, noise immunity, availability on the majority of flight routes and high accuracy that depends mostly on cartographic data require further development of data fusion techniques for processing information in SINS+CENS integration scheme.

Principle of action of CENS is based on determining the current object position by correlative comparison the measured realization of geophysical field with template one and finding the best (extreme) match [2]. One of data processing stages in CENS is forming the informative features of geophysical field and their selection.

The field sensor in general case outputs data that requires preprocessing, in particular extraction of informative features and transformation to the form represented in cartographic information. Detection of characteristic features of the field in general case is determined in feature space, that contains all possible features of the field. The features of optical field of the ground surface can be classified by the following types:

– geometrical (feature points, contours or binary large areas – BLOB or Binary Large OBjects);

– spectral (spectral descriptions on the basis of Fourier or wavelet transforms);

– structural (on the basis of morphological analysis);

– statistical (on the basis of probability distribution, method of histograms, etc).

II. PROBLEM STATEMENT

The problem statement is therefore formulated as following. It is necessary to select such components in feature vector which provide relevant representation of field realization in three-dimensional space. Relevance and completeness of selected features will be determined based on maximum entropy principle [3].

If the decision to match the current realization with definite template is made based on the incomplete information then it must be selected from the probability distribution which maximizes entropy under the given constraints on the distribution.

To maximize entropy it is clear that the more features are included in the feature vector the better relevance of made decision will be. However, there must be compromise between dimension of feature vector and available computing resources of airborne computer. Let's formulate criteria of feature selection which can be implemented at the stage of cartographic data preparation. Feature vector in general case will contain the following data:

$$\mathbf{X} = \Bigg[\underbrace{\mathbf{x}_1 = \begin{Bmatrix} x_1.\text{field}_1 \\ \vdots \\ x_1.\text{field}_{N1} \end{Bmatrix} \mathbf{x}_2 = \begin{Bmatrix} x_2.\text{field}_1 \\ \vdots \\ x_2.\text{field}_{N2} \end{Bmatrix} \ldots}_{\text{class 1}}$$

$$\ldots \underbrace{\mathbf{x}_p = \begin{Bmatrix} x_p.\text{field}_1 \\ \vdots \\ x_p.\text{field}_{Np} \end{Bmatrix} \mathbf{x}_{p+1} = \begin{Bmatrix} x_{p+1}.\text{field}_1 \\ \vdots \\ x_{p+1}.\text{field}_{Np+1} \end{Bmatrix} \ldots}_{\text{class } K} \Bigg]$$

where each class corresponds to the definite features with numerical and logical values. For example, for class of Speed-Up Robust Feature (SURF) feature points there are the following numerical values: descriptor itself by 8×8 dimension, descriptor orientation, Laplacian, point coordinates on image, etc.

It is necessary to select class of features and to determine the minimal number of features to represent the given template without loss of matching reliability. It can be estimated by probability of false matching of current ith feature vector with jth template. Let's minimize the probability of false matching, since maximization of probability of true matching can be done in expense of unlimited increase of feature vector size.

$$\mathbf{X} = \arg\min_{K,N} P\left(\mathbf{X}_i = \mathbf{X}_j\right), \quad i \neq j. \qquad (1)$$

Further consideration will be done on the example of feature points taken from image by SURF method [4].

III. Problem Analysis

Many statistical methods for evaluating the worth of feature subsets based on characteristics of the training data are only applicable to numeric features. Furthermore, these measures are often monotonic (increasing the size of the feature subset can never decrease performance).

In [5] a feature selection algorithm based on ideas from information theory and probabilistic reasoning is introduced. The rationale behind their approach is that, since the goal of an induction algorithm is to estimate the probability distributions over the class values, given the original feature set, feature subset selection should attempt to remain as close to these original distributions as possible. One problem with the algorithm is that it requires features with more than two values to be encoded as binary in order to avoid the bias that entropic measures have toward features with many values. This can greatly increase the number of features in the original data, as well as introducing further dependencies.

Feature weighting can be viewed as a generalization of feature selection. In feature selection, feature weights are restricted to 0 or 1 (a feature is used or it is not). Feature weighting allows finer differentiation between features by assigning each a continuous valued weight. Algorithms such as nearest neighbour (that normally treat each feature equally) can be easily modified to include feature weighting when calculating similarity between templates. One thing to note is that, in general, feature weighting algorithms do not reduce the dimensionality of the data. Unless features with very low weight are removed from the data initially, it is assumed that each feature is useful for induction; its degree of usefulness is reflected in the magnitude of its weight.

IV. Problem Solution

Main difference of proposed algorithm is to use the correlation coefficients as a measure of feature relevance. Mutual correlation between features allows us to eliminate one of them without loss of matching reliability. Let's introduce the measure of feature similarity inside a template as following:

$$M_{\mathbf{X}} = \frac{N\overline{r_{ii}}}{\sqrt{N + N(N-1)\overline{r_{ij}}}}, \qquad (2)$$

where N is the number of features of the same class in a template, $\overline{r_{ij}}$ is the averaged value of cross-correlation between ith and jth features, $\overline{r_{ii}}$ is full correlation coefficient of template features.

Developed measure is normalized, it reaches to minimal value – zero under the absence of components in feature vector ($N = 0$) and tends to one in asymptote under unlimited increase of N.

General algorithm of forming and selecting of features is performed at the stage of cartographic data preparation and is the following.

The template is processed by definite method of feature detection. Let it will be SURF method. In general case the descriptor of feature point includes the following information: coordinates $P=\{x,y\}$, scale of Gaussian filter $M=\{\sigma\}$, gradient orientation $R=\{\varphi\}$, 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.

Then according to (2) the normalized correlation matrix **K** is calculated with elements:

$$K_{ij}=\frac{D_i(64)D_j(64)}{\sqrt{D_i^2(64)D_j^2(64)}}, \quad (3)$$

where indexes i, j correspond to the feature points on the same template and current images.

The descriptors are formed as matrix **D** by size 64×*N*, or 128×*N*, where *N* is a number of feature points. The function (3) can be found by *single* multiplication of matrix on its transpose version as following:

$$\mathbf{K}=\mathbf{D}^{\mathrm{T}}\cdot\mathbf{D} =$$

$$=\begin{bmatrix} D_1(1) & \cdots & D_1(N) \\ \vdots & \ddots & \vdots \\ D_{64}(1) & \cdots & D_{64}(N) \end{bmatrix}^{\mathrm{T}} \cdot \begin{bmatrix} D_1'(1) & \cdots & D_1'(N) \\ \vdots & \ddots & \vdots \\ D_{64}'(1) & \cdots & D_{64}'(N) \end{bmatrix}. \quad (4)$$

Since descriptor matrix is already normalized due to peculiarities of SURF method, then it is possible to state that each component K_{ij} of matrix (4) corresponds to (3).

Features with value of cross-correlation K_{ij}, $i \neq j$ greater than 0.98 are excluded from template.

Then the new feature vector $\mathbf{X}_{\text{basic}}$ is formed that contains all *N* relevant features with initial weights $w_i, i=1...N$ equal to ones.

Along with this, for unambiguous matching of a template it is necessary to have all features in measured signal which leads to 100% probability of recognition. Therefore, it is stated that feature vector with weighted components has physical sense of probability distribution function $p(\mathbf{Z}|\mathbf{X}_i)$ and requires the weight normalization by the following formula:

$$\frac{1}{N}\sum_{i=1}^{N} w_i = 1$$

Then the template field realization is subjected to planned transformations and distortions like perspective transforms, change of illumination, noise disturbances, etc. Also it may be the different images of the same area of ground surface taken in different day or season times.

Let's designate the total number of experiments for a single template as *K*. Then the relevance measure of each feature is estimated iteratively in decision making of matching by the following expression:

$$w_{i,k+1}=w_{i,k}\pm\Delta w, \quad k=1...K, \quad (5)$$

where addition corresponds to the case when *i*th feature has taken the participation in correct matching and correspondingly the minus sign will be if there has been no matching.

Degree of feature participation in decision making is assumed to be in the form of correlation coefficient $r_{ij}, j=1,...,M$, where *M* is the number of features found in *k* experiment.

After each iteration the weights are recalculated to normalize as

$$w_{i,k}=\frac{w_{i,k}}{\sum_{i=1}^{N} w_{i,k}}.$$

Experimental researches have been done on template images taken from quadrocopter DJI Phantom 3 Professional at Kiev region at GPS altitude 500. Realization of SURF method is used based on available code listing in Matlab by [6].

Template images have the resolution 540×960 px, mostly contain urban areas with informative regions. Template features are correlated with each other on preliminary stage. Features with value of cross-correlation K_{ij}, $i \neq j$ greater than 0.98 are excluded from template.

Decreasing of template size under threshold value T_{SURF} of SURF detector 0.001 is from 530 to 475 feature points, for T_{SURF} = 0.0005 the minimization is from 1082 to 976, and for T_{SURF} = 0.0001 it is from 2567 to 2311.

After decreasing the template size the experiments have been done to match the distorted fragment of template with several templates, including the resampling one. Results of experiments are represented in Table I.

As it can be seen the results of matching of feature points are satisfied in comparison with other incorrect templates, where the numbers of found matches (K_{ij} > 0.98) are less than threshold.

Moreover, the full template with high SURF threshold gives many similar feature points and therefore results in false matching of fragment as it is shown in Fig. 1.

Weighting of remaining features is done by setting the planned transformations of template that are actually taken as series of frames from onboard video. Trajectory of the whole flight is shown in Fig. 2.

TABLE I

RESULTS OF MATCHING BETWEEN FULL AND RESAMPLING TEMPLATE AND THEIR FRAGMENT

Rotation angle of fragment, deg	**3**	**5**	**8**	**10**	**15**
Ratio of correct matches for full and resampling template, T_{SURF} = 0.001	59/54	58/54	52/49	48/42	13/12
Ratio of correct matches for full and resampling template, T_{SURF} = 0.0005	128/114	130/119	121/112	98/90	25/20
Ratio of correct matches for full and resampling template, T_{SURF} = 0.0001	285/286	286/268	270/251	222/203	65/59

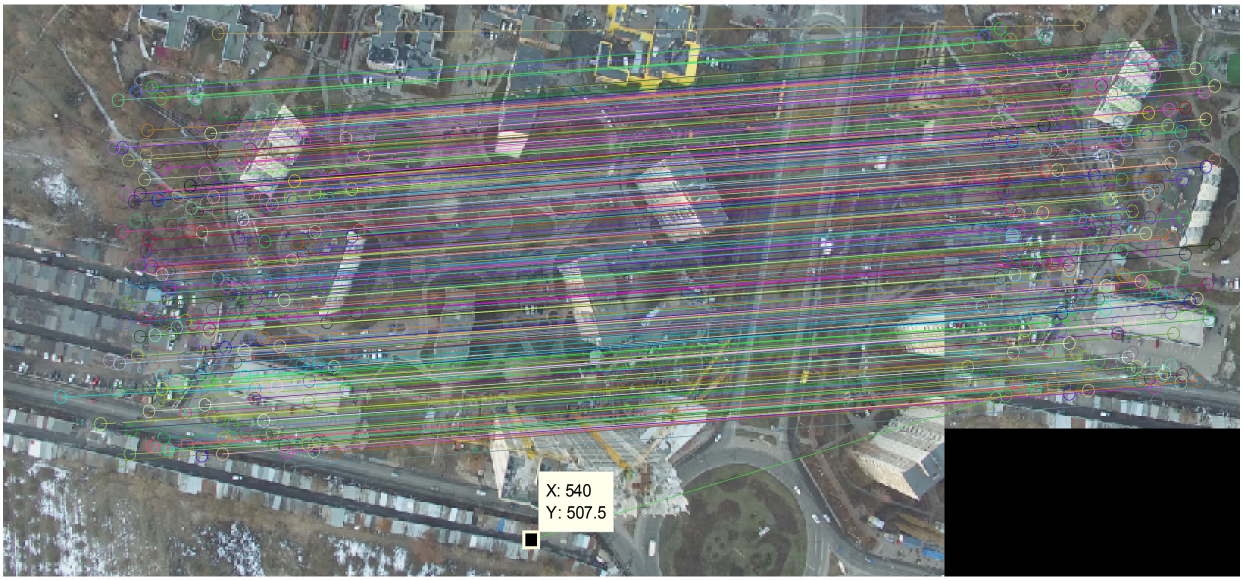

Fig. 1. Matching results of template with its fragment with false point matching due to redundancy of template features

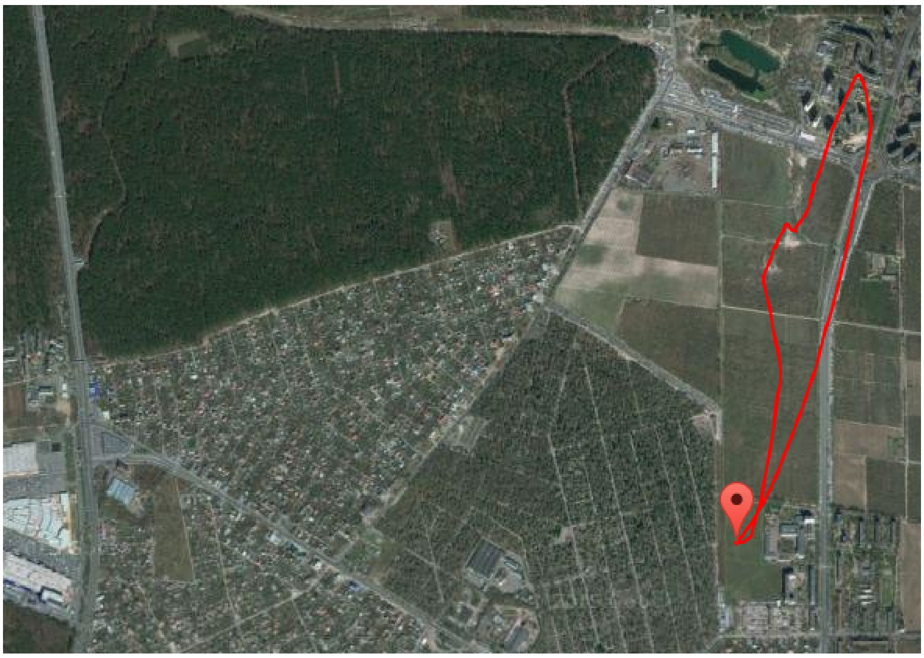

Fig. 2. Trajectory of flight to take series of template data

V. CONCLUSIONS

Thus, during experimental researches there was decreasing of template size without lose of matching accuracy up to 90% in average from total number of template features. Number of correct matching feature therefore also decreases to 9.49%. The obvious advantages of proposed algorithm are minimization of template dimensions, saving time of correlation matching and increasing of computing efficiency of CENS.

REFERENCES

[1] M.P. Mukhina. "Prospects of development correlation-extreme modern navigation systems". *AVIA-2013: Materials of the XI International scientific technical conference*. Kyiv: NAU, 2013. pp. 22.1–22.4.

[2] V. K. Baklitskiy. *Correlation-extreme methods of navigation and targeting*. Tver.: TO "Book Club", 2009, 360 p

[3] Eyink G. L., Kim S. A maximum entropy method for particle filtering //Journal of statistical physics. – 2006. – vol. 123. no. 5. pp. 1071–1128.

[4] Herbert Bay, Tinne Tuytelaars, Luc Van Gool. SURF: Speeded Up Robust Features. – Katholieke Universiteit Leuven. 2007. 13 p

[5] Koller D., Sahami M. Toward optimal feature selection. – Machine Learning: Proceedings of the 13-th International Conference on Machine Learning. 1996.

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

Received November 12, 2015

Mukhina Maryna. Ph.D. Associate professor.
Educational-Scientific Institute of Information Diagnostic Systems, National Aviation University, Kyiv, Ukraine.
Education: National Aviation University, Kyiv, Ukraine (2002).
Research area: images processing, navigation and control.
Publications: 33.
E-mail: m_mukhina@inbox.ru

Prymak Artem. 6th year of study student.
National Aviation University, Kyiv, Ukraine.
Education: National Aviation University, Kyiv, Ukraine, Bachelor (2014).
Research interests: the extremal control systems.
Publications: 3.
E-mail: primak.artem@gmail.com

Babeniuk Anna. 6th year of study student.
National Aviation University, Kyiv, Ukraine.
Education: National Aviation University, Kyiv, Ukraine, Bachelor (2014).
Research interests: correlation-extreme navigation system.
Publications: 5.
E-mail: anitabaly@mail.ru

М. П. Мухіна, А. П. Примак, А. М. Бабенюк. Алгоритм формування та вибору інформативних ознак в базі даних кореляційно-екстремальних навігаційних систем

Запропоновано ієрархічний підхід до сумісної обробки даних комплексної навігаційної системи. Нижній рівень обробки даних в кореляційно-екстремальній навігаційній системі включає формування та відбір найбільш інформативних ознак, таким чином розроблено розширений алгоритм на основі зваженої кореляції та вибірки ознак з еталону. Практичну ефективність алгоритму досліджено на основі використання методу виявлення характерних точкових ознак Speed-Up Robust Feature та використання його дискриптора. Спостерігалась мінімізація кількості особливих точок в еталоні без зниження якості співставлення нарівні зі зменшенням обчислювальних витрат.

Ключові слова: вибір інформативних ознак; метод максимальної ентропії; кореляційне зважування; інформативність еталону.

Мухіна Марина Петрівна. Кандидат технічних наук. Доцент.
Навчально-науковий інститут інформаційно-діагностичних систем, Національний авіаційний університет, Київ, Україна.
Освіта: Національний авіаційний університет, Київ, Україна, (2002).
Напрямок наукової діяльності: обробка зображень, навігація та керування.
Кількість публікацій: 33.
E-mail: m_mukhina@inbox.ru

Примак Артем Петрович. Студент 6-го курсу.
Національний авіаційний університет, Київ, Україна.
Освіта: Національний авіаційний університет, Київ, Україна, Бакалавр (2014).
Напрямок наукової діяльності: екстремальні системи управління.
Кількість публікацій: 3.
E-mail: primak.artem@gmail.com

Бабенюк Анна Миколаївна. Студентка 6-го курсу.
Національний авіаційний університет, Київ, Україна.
Освіта: Національний авіаційний університет, Київ, Україна, Бакалавр (2014).
Напрямок наукової діяльності: кореляційні-екстремальні навігаційні системи.
Кількість публікацій: 5.
E-mail: anitabaly@mail.ru

М. П. Мухина, А. П. Примак, А. Н. Бабенюк. Алгоритм формирования и выбора информативных признаков в базе данных корреляционно-экстремальных навигационных систем

Предлагается иерархический подход к совместной обработке данных в комплексной навигационной системе. Нижний уровень обработки данных в корреляционно-экстремальной навигационной системе включает формирование и отбор наиболее информативных признаков, таким образом, разработан расширенный алгоритм на основе взвешенной корреляции и выборки признаков из эталона. Практическая эффективность алгоритма исследована на основе использования метода обнаружения характерных точечных признаков Speed-Up Robust Feature и использования его дескриптора. Наблюдалась минимизация количества особенных точек в эталоне без снижения качества сопоставления наряду с уменьшением вычислительных затрат.

Ключевые слова: выбор информативных признаков; метод максимальной энтропии; корреляционное взвешивание; информативность эталона.

Мухина Марина Петровна. Кандидат технический наук, доцент.
Учебно-научный институт информационно-диагностических систем, Национальный Авиационный Университет, Киев,Украина.
Образование: Национальный авиационный универститет, Киев, Украина, (2002).
Направление научной деятельности: обработка изображений, навигация и управление.
Количество публикаций: 33.
E-mail: m_mukhina@inbox.ru

Примак Артем Петрович. Студент 6-го курса.
Национальный авиационный университет, Киев, Украина.
Образование: Национальный авиационный университет, Киев, Украина, Бакалавр (2014).
Направление научной деятельности: экстремальные системы управления.
Количество публикаций: 3.
E-mail: primak.artem@gmail.com

Бабенюк Анна Николаивна. Студентка 6-го курса.
Национальный авиационный университет, Киев, Украина.
Образование: Национальный авиационный университет, Киев, Украина, Бакалавр (2014).
Направление научной деятельности: корреляционные-экстремальные навигационные системы.
Количество публикаций: 5.
E-mail: anitabaly@mail.ru