

УДК 004.932

## СОЗДАНИЕ ПАНОРАМНЫХ АЭРОФОТОСНИМКОВ НА ОСНОВЕ МНОГОПОЛОСНОГО СМЕШИВАНИЯ\*

М. Н. Фаворская, А. И. Пахирка, А. Г. Зотин, В. В. Буряченко

*Сибирский государственный университет  
науки и технологий им. академика М. Ф. Решетнева,  
660037, г. Красноярск, просп. им. газеты «Красноярский рабочий», 31  
E-mail: favorskaya@sibsau.ru*

Панорамный аэрофотоснимок, составленный из набора последовательных изображений, полученных с беспилотного летательного аппарата, позволяет более точно распознавать визуальные объекты. Автоматическая сшивка аэрофотоснимков представляет собой сложную техническую задачу, особенно при наличии артефактов регистрирующей аппаратуры и условий съёмки. Считается, что последовательность сшиваемых изображений известна и определяется полётной траекторией. Предлагается метод сшивки аэрофотоснимков на основе выравнивания освещённости и многополосного смешивания. Исследованы возможности применения точечных дескрипторов SURF, FAST, FAST-ER и KAZE для сшивки изображений. Показано, что точечные дескрипторы на основе нелинейной диффузной фильтрации дают лучшие результаты по точности при наличии глобальных и/или локальных визуальных искажений сшиваемых изображений.

*Ключевые слова:* аэрофотоснимок, панорама, точечные дескрипторы, выравнивание освещённости, многополосное смешивание.

DOI: 10.15372/AUT20180303

**Введение.** Видеосъёмка, осуществляемая с использованием беспилотных летательных аппаратов (БПЛА), имеет огромное значение во многих отраслях промышленности и сельского хозяйства как дешёвый и адаптируемый к погодным условиям способ мониторинга локальных территорий [1]. Модельный ряд БПЛА включает широкий класс устройств с различными лётными и нагрузочными характеристиками. Чем ниже такие характеристики, тем более БПЛА подвержен влиянию воздушных потоков, вызывающих нестационарность изображений, визуальные искажения, наличие размытых кадров. Одним из основных ограничений применения БПЛА является относительно небольшой охват территории. Создание панорамных снимков позволяет отчасти решить данную задачу.

В общем случае создание панорамного снимка требует точного выравнивания изображений в заданной системе координат с последующим использованием гомографических преобразований, последовательно применяемых к паре сшиваемых снимков [2]. Будем считать, что сведения о траектории полёта БПЛА известны, а пользователь задаёт кадры начала и окончания видеопоследовательности для формирования панорамного снимка. Исходя из указанных данных, можно построить связный граф, в котором вершинами являются кадры, а рёбра, соединяющие их, формируются в соответствии с полётной траекторией [3]. Такая процедура позволяет избавиться от дорогостоящего сопоставления случайных пар неотсортированных изображений. Примем, что центральное изображение для панорамного снимка устанавливает пользователь, ориентируясь на объекты интереса, а изображения отображены по критериям качества цветности и контраста, равномерности освещённости, отсутствия размытия.

\*Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 16-07-00121а).

Совмещение и выравнивание изображений, которые могут быть получены при различных угловых положениях камеры в 3D-пространстве, являются основными процедурами создания панорамного снимка. Известны два подхода выравнивания изображений с использованием регионов и точечных дескрипторов. Считается, что первый подход обеспечивает большую точность совмещения [4], однако имеет ограничения по параметрам съёмки (они должны быть близкими к параметрам съёмки исходного снимка). Такой подход не рекомендуется для сшивки большого количества изображений. Второй подход на основе разнообразных точечных дескрипторов используется более широко [5]. При правильном подборе кадров можно избежать проблем с низкой контрастностью, размытостью или наличием слабо текстурированных областей, которые отрицательно влияют на функционирование точечных дескрипторов.

В данной работе исследуются четыре типа точечных дескрипторов на предмет точного совмещения изображений. Также предлагается метод сшивки аэрофотоснимков на основе выравнивания освещённости и многополосного смешивания.

**Анализ точечных дескрипторов.** Рассмотрим работу используемых дескрипторов. Быстрые дескрипторы FAST (Features from Accelerated Segment Test) и FAST-ER (FAST Enhanced Repeatability) предназначены для обнаружения углов, дескриптор SURF (Speeded-Up Robust Feature) учитывает однородное масштабирование пространства, а дескриптор KAZE (в переводе с японского «ветер») показывает хорошие результаты при неоднородном масштабном изменении, когда сшиваемые изображения имеют разные углы наклона с расхождением значений до 3–5°.

Является ли центральный пиксел  $p$  в дескрипторе FAST углом, проверяется сравнением значения его яркости со значениями яркостей 16 пикселей окружения  $\{q_1, q_2, \dots, q_{16}\}$  с радиусом 3 (рис. 1, *c*). Для ускорения работы алгоритма вначале рассматриваются пиксели с номерами 1, 5, 9 и 13. Если, по крайней мере, три пикселя из четырёх удовлетворяют пороговому критерию, то пиксел  $p$  считается углом. Если, как минимум, три значения из четырёх находятся ниже порогового значения  $I_p + T_{\text{FAST}}$ , то пиксел  $p$  не является точкой интереса и отбрасывается. Такой быстрый дескриптор имеет ряд недостатков, связанных с тем, что может быть найдено несколько соседних пикселей, интерпретируемых как угол, а определение яркости угла (светлый или тёмный) требует дополнительных действий. В связи с этим в [6] предложили использовать методы машинного обучения для построения правил на основе дерева решений. Поскольку процедуру вычисления дескриптора FAST можно представить в виде троичного дерева решений, в [6] предложено обобщение дескриптора, названного FAST-ER (анализируется 48 пикселей окружения). При этом де-

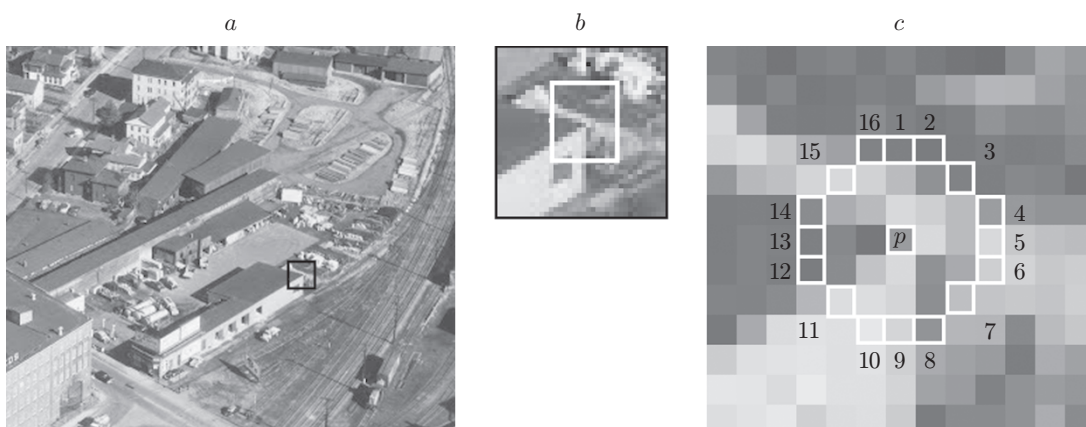


Рис. 1. Нахождение углов: *a* — исходное изображение, *b* — увеличенный фрагмент, *c* — уголкового дескриптора FAST

рево решений оптимизируется с использованием метода отжига, что делает дескриптор FAST-ER инвариантным к вращению, отражению и инверсии интенсивности.

Для идентификации особой точки в дескрипторе SURF строится массив из 64 (в расширенной версии 128) чисел. Размер окна для поиска дескриптора равен  $20 \text{ scale}$ , где  $\text{scale}$  — масштаб, в котором была найдена особая точка. Дескриптор содержит описание градиентов для 16 квадрантов вокруг особой точки. Далее квадрант разбивается на 16 более мелких квадрантов. В каждом квадранте берётся регулярная сетка  $5 \times 5$  и для каждой точки сетки ищется градиент с помощью фильтра Хаара. Размер фильтра Хаара равен  $2 \text{ scale}$  и для первой октавы составляет  $4 \times 4$ . Следует отметить, что при расчёте фильтра Хаара изображение не поворачивается, фильтр применяется в обычных координатах изображения. Затем полученные координаты градиента ( $dX, dY$ ) поворачиваются на угол, соответствующий ориентации квадранта. Также для описания точки используется знак следа матрицы Гессе. Для светлых точек на тёмном фоне след отрицателен, а для тёмных точек на светлом фоне след положителен. Таким образом, дескриптор SURF различает светлые и тёмные пятна.

В ряде дескрипторов используется гауссово размытие для удаления шума, однако при этом размываются контуры изображения, что является нежелательным следствием. Для сохранения контуров были предложены дескрипторы, основанные на построении нелинейной диффузной фильтрации [7]. Ядром дескриптора KAZE выступает аппроксимация классической нелинейной диффузной фильтрации, которая описывается функцией потока в виде нелинейных уравнений в частных производных:

$$\frac{\partial L}{\partial t} = \text{div}(c(x, y, t)\nabla L), \quad (1)$$

где  $L$  — функция потока;  $\text{div}$  — оператор дивергенции;  $\nabla$  — градиентный оператор;  $c$  — функция проводимости (может быть скалярной функцией или тензором);  $t$  — время как параметр масштаба.

В задачах компьютерного зрения функция проводимости  $c$  в выражении (1) рассчитывается через величину градиента или аппроксимируется численными методами. В последнем случае нелинейная диффузная фильтрация представляется в виде аддитивного оператора разделения (AOS — Additive Operator Splitting). Схема AOS является дискретным вариантом линейной неявной схемы. Используя данную схему, можно получать устойчивые нелинейные масштабные пространства для любого шага очень эффективным способом, а именно, решая систему трёхдиагональных линейных уравнений с помощью алгоритма Томаса. Алгоритм Томаса является упрощением метода исключения Гаусса, когда решение достигается за  $O(n)$  операций вместо  $O(n^3)$  операций. По сути, дальнейшее построение дескриптора KAZE аналогично построению модифицированного дескриптора SURF. Время вычисления дескриптора KAZE сравнимо со временем вычисления дескриптора SIFT (Scale-Invariant Feature Transform), причём более половины этого времени требуется для формирования нелинейного масштабного пространства. В [8] предложен ускоренный дескриптор KAZE, использующий преимущества явной и линейной неявной диффузии.

На рис. 2 представлен пример обнаружения точечных особенностей на изображении с применением дескрипторов FAST, FAST-ER, SURF и KAZE.

**Совмещение точечных дескрипторов.** Один из самых распространённых способов оценки параметров модели — алгоритм RANSAC (RANdom SAMple Consensus) [9], концепция которого основана на разделении исходных данных на выбросы и попадания. Выбросами являются ложные точки, шумы, случайные включения, т. е. такие точки, на основании которых нельзя построить требуемую модель. Попаданиями называются точки, удовлетворяющие модели. Этот алгоритм выполняется итеративно. У алгоритма RANSAC имеется ряд недостатков, которые иногда не позволяют получить требуемого результата

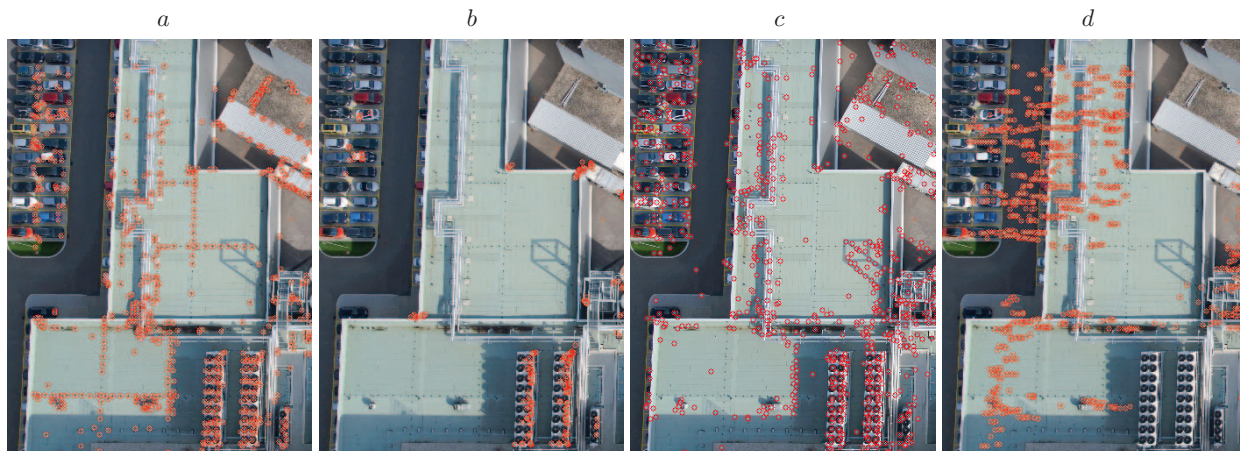


Рис. 2. Нахождение точечных особенностей на изображении EP-00-00012\_0119\_0064.jpg из набора данных Swiss dataset (Le Bourget Airport) [10]: *a* — дескриптор FAST, *b* — дескриптор FAST-ER, *c* — дескриптор SURF, *d* — дескриптор KAZE

за приемлемое время, в частности отсутствие верхней границы времени вычисления параметров модели. В настоящее время разработаны, реализованы и протестированы модификации алгоритма. Одной из таких модификаций является алгоритм MLESAC (Maximum Likelihood Estimation Sample Consensus) [11], который вычисляет вероятность гипотезы сопоставления точечных особенностей, представляя распространение ошибки как смешивание моделей. Для одновременного вычисления параметров модели и доли выбросов применяется схема максимизации ожидания. Для сшивки изображений требуется подобрать четыре наилучших сопоставления точечных особенностей, координаты которых используются для вычисления коэффициентов матрицы гомографии

$$\begin{bmatrix} x_j \\ y_j \\ 1 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}, \quad (2)$$

где  $h_1, \dots, h_9$  — коэффициенты матрицы;  $(x_i, y_i)$  и  $(x_j, y_j)$  — координаты соответствующих точек.

При проективном преобразовании  $h_9 = 1$  выражение (2) принимает вид

$$x_j = \frac{h_1 x_i + h_2 y_i + h_3}{h_7 x_i + h_8 y_i + 1}, \quad y_j = \frac{h_4 x_i + h_5 y_i + h_6}{h_7 x_i + h_8 y_i + 1}.$$

Тогда пара соответствующих точек связана соотношениями

$$\begin{aligned} h_1 x_i + h_2 y_i + h_3 - h_7 x_i x_j - h_8 x_j y_i &= x_j; \\ h_4 x_i + h_5 y_i + h_6 - h_7 x_i y_j - h_8 y_i y_j &= y_j. \end{aligned} \quad (3)$$

Решая систему уравнений (3) для четырёх лучших соответствий, найденных по алгоритму RANSAC или MLESAC, можно пересчитать координаты сшиваемых изображений в перекрывающейся области.

**Бесшовная стыковка изображений.** Бесшовное смешивание изображений, как правило, является необходимым после построения панорамного снимка в связи с возникающими локальными артефактами освещённости, рассогласования текстурной мозаики и т. д.

Этот этап включает компенсацию освещённости с последующим применением алгоритма смешивания в перекрывающихся областях двух сшиваемых изображений. Компенсация освещённости достигается вычислением коэффициентов усиления  $g$  в выражении, минимизирующем сумму квадратов разностей освещённости в перекрывающихся областях двух изображений:

$$e = \frac{(g_a \bar{I}_{ab} - g_b \bar{I}_{ba})^2}{\sigma_I^2} + \frac{(1 - g_a)^2}{\sigma_g^2}, \quad (4)$$

где  $\bar{I}_{ab}$  и  $\bar{I}_{ba}$  — средние значения интенсивностей в перекрывающихся областях изображений  $a$  и  $b$ ;  $\sigma_I$  — стандартное отклонение ошибки нормализованной интенсивности ( $\sigma_I = 10,0$ );  $\sigma_g$  — стандартное отклонение ошибки коэффициента усиления  $g$  ( $\sigma_g = 0,1$ ).

Примем, что интенсивность изображения  $a$  не требует изменений. Второе слагаемое в выражении (4) введено для того, чтобы коэффициент усиления  $g_a \rightarrow 1$ . Тогда целевая функция (4) минимизируется относительно коэффициентов усиления взятием производной и приравнованием её к нулю. Отметим, что существуют и другие способы компенсации освещённости, например метод  $\gamma$ -коррекции.

Среди известных методов сшивки панорамных изображений [12] многополосное смешивание считается самым эффективным, но наиболее затратным методом. Идея метода заключается в смешивании низких частот на большом пространственном диапазоне, а высоких частот — на малом пространственном диапазоне [13].

Алгоритм многополосного смешивания является рекурсивным. Обозначим весовую функцию в евклидовом пространстве как  $W(x, y) = w(x)w(y)$ , где функции  $w(x)$  и  $w(y)$  изменяются линейно в диапазоне от 0 на границах и до 1 в центре изображения. Вначале находим максимальную весовую функцию  $W_{\max}^a(x, y)$ , которая определяется по множествам пикселей  $a$  и  $b$  и удовлетворяет условию

$$W_{\max}^a(x, y) = \begin{cases} 1, & \text{если } W^a(x, y) = \arg \max_b W^b(x, y), \\ 0 & \text{в противном случае.} \end{cases} \quad (5)$$

Далее для каждого поддиапазона вычисляется функция пространственных частот  $B_\sigma^a(x, y)$  и весовая функция  $W_\sigma^a(x, y)$  в диапазоне длин волн  $\lambda \in [0, \sigma]$ . Для нулевого шага эти функции имеют вид

$$\begin{aligned} B_\sigma^a(x, y) &= I^a(x, y) - I_\sigma^a(x, y); \\ I_\sigma^a(x, y) &= I^a(x, y) * g_\sigma(x, y); \\ W_\sigma^a(x, y) &= W_{\max}^a(x, y) * g_\sigma(x, y), \end{aligned} \quad (6)$$

где  $I^a(x, y)$  — изображение в текущем диапазоне;  $I_\sigma^a(x, y)$  — свёртка изображения с гауссианом;  $g_\sigma(x, y)$  — фильтр Гаусса со стандартным отклонением  $\sigma$ .

На нулевом шаге выражения (6) позволяют выявить высокочастотные составляющие сшиваемых изображений. На последующих шагах ( $k \geq 1$ ) функции  $B_{(k+1)\sigma}^a(x, y)$  и  $W_{(k+1)\sigma}^a(x, y)$  вычисляются с бóльшим значением стандартного отклонения:

$$\begin{aligned} B_{(k+1)\sigma}^a(x, y) &= I_{k\sigma}^a(x, y) - I_{(k+1)\sigma}^a(x, y), \\ I_{(k+1)\sigma}^a(x, y) &= I_{k\sigma}^a(x, y) * g_{\sigma'}(x, y), \\ W_{(k+1)\sigma}^a(x, y) &= W_{k\sigma}^a(x, y) * g_{\sigma'}(x, y), \end{aligned}$$

где  $\sigma'$  — стандартное отклонение в последующих поддиапазонах,  $\sigma' = (2k + 1)^{1/2}\sigma$ .

В результате высокочастотные составляющие размываются на малом пространственном диапазоне, а низкочастотные — на большом пространственном диапазоне. Интенсивность полосы смешивания рассчитывается как средневзвешенное значение по формуле

$$I_{k\sigma}^{\text{multi}}(x, y) = \sum_{i=1}^N B_{k\sigma}^a(x, y) W_{k\sigma}^a(x, y) / \sum_{i=1}^N W_{k\sigma}^a(x, y), \quad (7)$$

где  $N$  — количество поддиапазонов.

Таким образом, результатом многополосного смешивания является суммирование изображений по всем поддиапазнам.

**Экспериментальные исследования.** При проведении экспериментов использовался набор изображений Swiss dataset (Le Bourget Airport) [10], полученный с помощью дро-



Рис. 3. Панорамный аэрофотоснимок, полученный на основе многополосного смешивания из набора деформированных изображений с использованием дескрипторов:  $a$  — SURF,  $b$  — KAZE

**Время работы алгоритмов сопоставления изображений**

Разрешение, пиксели	SURF, мс		FAST-ER, мс		FAST, мс		KAZE, мс	
	D <sup>1</sup>	R/M <sup>2</sup>	D	R/M	D	R/M	D	R/M
640 × 360	535	8/9	22	4/5	13	6/7	288	6/7
852 × 480	782	12/13	33	5/6	19	7/8	495	7/8
1280 × 720	1380	14/16	65	5/6	29	7/8	1107	8/9
1600 × 900	1965	15/18	90	6/7	34	7/9	1686	8/10
1920 × 1080	2844	17/18	132	6/8	43	8/9	2415	11/13

*Примечания:* 1) D — время нахождения дескриптора; 2) R/M — время сопоставления точечных особенностей по алгоритмам RANSAC/MLESAC.

на eBee RTK, имеющего следующие характеристики: масштаб 3,14 см/пиксел, область 0,15 км<sup>2</sup>, высота полёта 120 м, число изображений 128. Примеры панорамных аэрофотоснимков на основе многополосного смешивания представлены на рис. 3. Как видно из рисунка, панорамный снимок с использованием дескрипторов SURF имеет определённые артефакты сшивки. Отметим, что аналогичные артефакты возникали и при сшивке изображений с применением дескрипторов FAST и FAST-ER. Сшивка на основе дескриптора KAZE обеспечивает лучшие визуальные результаты.

Тестирование проводилось на компьютере следующей конфигурации: процессор AMD Athlon II X3 3.2GHz, оперативная память 8GB DDR3 1333MHz, видеокарта NVIDIA GeForce GTX 650. В таблице представлено время работы алгоритма нахождения точечных особенностей и сопоставления изображений с разным разрешением на основе дескрипторов SURF, FAST, FAST-ER и KAZE.

Полученные данные показывают, что использование алгоритма RANSAC обеспечивает большую скорость сопоставления точечных особенностей в среднем на 15 % по сравнению с алгоритмом MLESAC. При этом алгоритм MLESAC повышает точность сопоставления в среднем на 5 %, достигая 81 % для исследуемого набора изображений.

**Заключение.** Применение дескрипторов FAST и FAST-ER позволяет организовать сопоставление изображений (с разрешением до 1920 × 1080 пикселей) для формирования панорамного снимка в реальном режиме времени. Однако наилучшие результаты с использованием технологии многополосного смешивания обеспечивает дескриптор KAZE, работающий по методу нелинейной диффузной фильтрации.

### СПИСОК ЛИТЕРАТУРЫ

1. Белоконь С. А., Золотухин Ю. Н., Нестеров А. А. Планирование маршрутов движения летательного аппарата с использованием гладких траекторий // Автометрия. 2017. **53**, № 1. С. 3–11.
2. Kang L., Wu L., Wei Y. et al. A highly accurate dense approach for homography estimation using modified differential evolution // Eng. Appl. Artificial Intell. 2014. **31**, N 4. P. 68–77.
3. Wang Z., Chen Y., Zhu Z., Zhao W. An automatic panoramic image mosaic method based on graph model // Multimedia Tools and Appl. 2015. **75**, N 5. P. 2725–2740.
4. Ghannam S., Abbott A. L. Cross correlation versus mutual information for image mosaicing // Intern. Journ. Adv. Comput. Sci. and Appl. 2013. **11**, N 4. P. 94–102.
5. Jain L. C., Favorskaya M., Novikov D. Panorama construction from multi-view cameras in outdoor scenes // Computer Vision in Control Systems-2 /Eds. M. N. Favorskaya, L. C. Jain. Switzerland: Springer, 2015. Vol. 75. P. 71–108.

6. **Rosten E., Porter R., Drummond T.** Faster and better: A machine learning approach to corner detection // IEEE Trans. Pattern Analys. and Mach. Intell. 2010. **1**, N 32. P. 105–119.
7. **Weickert J., ter Haar Romeny B., Viergever M. A.** Efficient and reliable schemes for non-linear diffusion filtering // IEEE Trans. Image Process. 1998. **7**, N 3. P. 398–410.
8. **Alcantarilla P. F., Nuevo J., Bartoli A.** Fast explicit diffusion for accelerated features in nonlinear scale spaces // Proc. of British Machine Vision Conf. (BMVC'2013). Bristol, UK, Sept. 9–13, 2013. P. 13.1–13.11.
9. **Fischler M. A., Bolles R. C.** Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography // Commun. ACM. 1981. **24**, N 6. P. 381–395.
10. **SenseFly.** Swiss dataset (Le Bourget Airport). URL: <https://www.sensefly.com/drones/example-datasets.html> (дата обращения: 6.01.2018).
11. **Torr P. H. S., Zisserman A.** MLESAC: A new robust estimator with application to estimating image geometry // Computer Vision and Image Understanding. 2000. **78**, N 1. P. 138–156.
12. **Favorskaya M. N., Buryachenko V. V., Zotin A. G., Pakhirka A. I.** Video completion in digital stabilization task using pseudo-panoramic technique // Proc. of the Intern. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Moscow, Russia, May 15–17, 2017. Vol. XLII-2/W4. P. 83–90.
13. **Brown M., Lowe D. G.** Automatic panoramic image stitching using invariant features // Intern. Journ. Comput. Vision. 2007. **74**, N 1. P. 59–73.

*Поступила в редакцию 16 января 2018 г.*

---