

ФИЛЬТРАЦИЯ ОШИБОЧНЫХ ПОКАДРОВЫХ РЕЗУЛЬТАТОВ  
В ПРОЦЕССЕ ЛОКАЛИЗАЦИИ ПРЯМОУГОЛЬНЫХ  
ПЛОСКИХ ОБЪЕКТОВ ПРИ ВИДЕОСЪЕМКЕ С ИСПОЛЬЗОВАНИЕМ  
СОГЛАСОВАНИЯ ГРАФА ПРЕОБРАЗОВАНИЙ

© 2020 г. Е. В. Емельянова<sup>1,2,\*</sup>, Б. И. Савельев<sup>2</sup>, К. Б. Булатов<sup>2,3</sup>

<sup>1</sup>Национальный исследовательский технологический университет “МИСиС”

119049 Москва, Ленинский проспект, д. 4, Россия

<sup>2</sup>ООО “Смарт Эндженс Сервис”

117312 Москва, проспект 60-летия Октября, д. 9, Россия

<sup>3</sup>Федеральный исследовательский центр “Информатика и управление” Российской академии наук  
127051 Москва, Большой Картеный переулок, д. 19, Россия

\*E-mail: e.emelianova@smartengines.ru

Поступила в редакцию 08.10.2019 г.

После доработки 21.10.2019 г.

Принята к публикации 29.10.2019 г.

Детектирование прямоугольных объектов является важной частью распознавания документов и автоматизации документооборота. При этом в процессе детектирования может появляться выбросовый шум – ложноопределенные координаты расположения документа, негативно влияющий на общее качество детектирования объекта и дальнейшей обработки. В работе поставлена задача фильтрации такого шума при детектировании плоского прямоугольного объекта на кадрах видеопоследовательности путем построения единой системы координат. Предлагаемый метод основывается на согласовании графа оцененных проективных преобразований кадров с шаблонами и кадрами между собой. В проведенных экспериментах использовался смоделированный выбросовый шум, применяющийся к некоторым кадрам видеопоследовательности. Для оценки качества работы алгоритма сравниваются результаты расположения объекта до и после модуля согласования графа с истинным расположением объекта. Данная оценка представлена на основе открытой базы данных MIDV-500 видеосъемки документов, удостоверяющих личность. На этом наборе данных продемонстрировано снижение накопленной ошибки по сравнению с результатами до использования согласования графа при вычислении проективных преобразований.

**Ключевые слова:** согласование графа преобразований, проективное преобразование, единая система координат, видеопоследовательность, детектирование объектов, накопленная ошибка

**DOI:** 10.31857/S0235009220010084

## ВВЕДЕНИЕ

Вследствие научно-технического прогресса человечество стало переходить на хранение документов в электронном виде и это затронуло многие области, такие как медицина (Novack, Lim, 2019), инженерия (Moreno-García et al., 2019), история (Babu, Soumya, 2019) и многие другие. Количество документов, которое необходимо перенести в электронный формат, неизбежно растет с каждым годом, из-за чего появляется актуальность в автоматизации процесса распознавания информации с документа, чтобы увеличить скорость работы и избавить людей от ручного труда по оцифровке данных, указанных в документе. Одним из главных этапов распознавания документа является локализация объекта на кадре. В

данном процессе могут возникать выбросы, т.е. расположение объекта распознается существенно неточно.

Цель настоящей работы – улучшение качества детектирования документа на видеопоследовательности по сравнению с результатами до использования согласования графа и для этого предлагается метод фильтрации выбросов в процессе локализации прямоугольных плоских объектов с использованием модуля согласования графа проективных преобразований.

В настоящее время существуют несколько методов детектирования плоских объектов в видео-потоке – оптический поток (Kelson et al., 2008), при котором строится векторное поле скоростей выделенных пикселей изображения; метод поис-



Рис. 1. Функциональная схема метода фильтрации.

ка особых точек (Turcot, Lowe, 2009), который основан на вычислении характерных особенностей на изображении и нахождении соответствия между ними и другие. Во всех этих методах имеется вероятность появления выбросового шума.

### ОПИСАНИЕ МЕТОДА ФИЛЬТРАЦИИ ВЫБРОСОВ С ИСПОЛЬЗОВАНИЕМ СОГЛАСОВАНИЯ ГРАФА ПРЕОБРАЗОВАНИЙ

Входными данными предложенного метода являются  $n$  кадров видеопоследовательности, на которых изображен плоский четырехугольный объект, где  $I_n$  – очередной монохромный кадр. Расположение объекта представлено в виде набора четырех упорядоченных двумерных векторов, содержащего в себе расположение углов документа. В представленном алгоритме существует две разновидности таких наборов: истинное расположение объекта на  $i$ -м кадре, которое будем называть эталоном  $q_i$ , и найденное расположение объекта на  $i$ -м кадре некоторым методом, описанным ниже, которое будем называть результатом детектирования  $g_i$ .

На рис. 1 представлена функциональная схема предложенного метода, где на вход алгоритма поступают кадры, которые приходят на два модуля:

– Модуль детектирования объекта на кадре, который осуществляет поиск образа плоского прямоугольного объекта с применением метода

one-shot learning (Skoryukina et al., 2018), при помощи поиска быстрого четырехугольника (Skogtukin et al., 2015). На выходе модуль имеет  $n$  результатов детектирования  $g_i$ .

– Модуль вычисления признаков, который осуществляет поиск свойств с использованием особых точек YACIPE (Lukoyanov et al., 2018) и дескрипторов RFD (Fan et al., 2014). На выходе модуль имеет базу признаков, которая содержит в себе особые признаки всех кадров.

Следующим этапом является построение графа матриц проективных преобразований (Хартсхорн, 1970). Вершинами графа являются кадры, ребрами – проективные преобразования между кадрами. Ребра строятся между соседними кадрами, следовательно, между ними осуществляется построение матриц проективных преобразований. Для поиска матриц проективных преобразований используется метод RANSAC (Martin, Robert, 1981).

Тогда координаты  $j$ -го кадра можно найти по формуле

$$g_j = H_{ij}g_i, \quad (1)$$

где  $H_{ij}$  – матрица проективного преобразования, переводящая набор точек  $i$ -го кадра в набор точек  $j$ -го кадра.

После этого к графу добавляется дополнительная нулевая вершина, которой не соответствует никакой кадр, ей соответствует шаблон докумен-

та. Она представляет собой координаты заданного прямоугольника. На рис. 2 представлен полученный граф, где более жирные ребра строятся на основе метода сопоставления особых точек, а менее жирные ребра, связывающие нулевую вершину с остальными, строятся на основе преобразования координат расположения углов документа.

Для обеспечения единой системы координат совокупности кадров достаточно задать граф проективных отображений, являющийся деревом. Для произвольного графа можно взять любое его оставное дерево, и оно будет обеспечивать единую систему координат. Построенный граф содержит циклы, т.е. не является деревом. В нем появляется избыточная информация, которая может содержать противоречия. В цикле композиция матриц проективных преобразований должна быть тождественно равна единичной матрице, и такой цикл называется согласованным. Из-за неточного построения матриц проективных преобразований будет расти накопленная ошибка.

Алгоритм согласования графа проективных преобразований (Савельев и др., 2018) заключается в следующем: метод SLAM (Newman, Ho, 2005) применяется для минимизации аккумулятивной ошибки, которая проявляется при замыкании цикла в графе проективных преобразований. Алгоритм согласования учитывает информацию со всех ребер графа, так как это обеспечивает лучший результат.

На данном этапе реализуется функция  $G(I_1, I_2, \dots, I_n, g_1, g_2, \dots, g_n)$  такая, чтобы минимизировался следующий функционал:

$$\frac{1}{n} \sum_{i=1}^n d(G(I_1, I_2, \dots, I_n, g_1, g_2, \dots, g_n), q_i) \rightarrow \min_g, \quad (2)$$

$$d(a, b) = \sqrt{\sum_{i=1}^4 \|a_i - b_i\|^2},$$

где  $a, b$  – некоторые четырехугольники.

На выходе метода согласования графа проективных преобразований имеются новые уточненные координаты расположения прямоугольного объекта на каждом кадре. Координаты до и после согласования графа будут отличаться, но, в случае с выбросом, они будут отличаться гораздо сильнее, чем при не выбросе. Для детектирования выбросовых кадров предлагается воспользоваться методом Иглевича и Хоглина (Iglewicz, Hoaglin, 1993). Для этого составляется множество расстояний между координатами расположения объекта на каждом кадре до и после согласования, определяется медиана  $M$  полученного множества, и медиана отклонения  $MAD$  от  $M$ . Выбросом считается тот кадр, разность расстояния и медианы  $M$

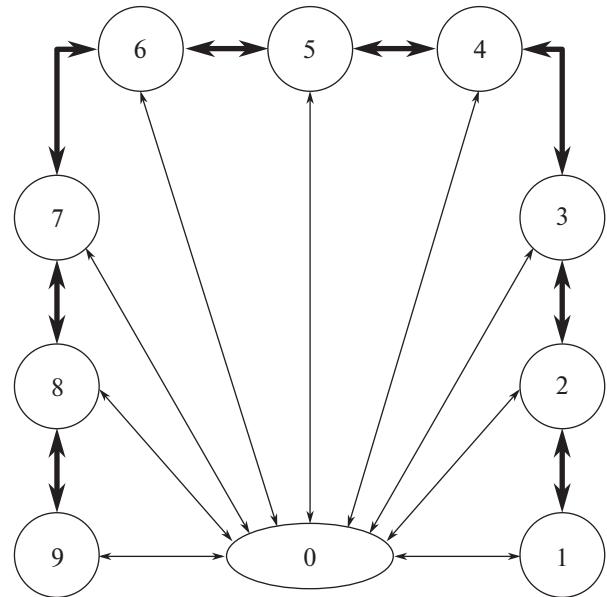


Рис. 2. Граф матриц проективных преобразований.

которого превышают значения двух медиан отклонений  $MAD$

$$dist_j = \frac{1}{4} \sum_{i=1}^4 \sqrt{|x_i - x'_i|^2 + |y_i - y'_i|^2}, \quad j = 1, \dots, n, \quad (3)$$

$$|dist_i - M| > 2 * MAD, \quad (4)$$

где  $M$  – среднее значение  $dist$ ,  $MAD$  – медиана отклонения от  $M$ .

В качестве результата используются откорректированные координаты расположения плоского объекта на кадре, которые были отмечены как выбросовые, полученные после согласования графа матриц проективных преобразований. Это улучшает качество детектирования объекта на всех кадрах видеопоследовательности. С кадрами, не отмеченными как выбросовые, ничего не происходит.

Для оценки качества работы алгоритма используется среднее значение разницы расстояний между результатом детектирования объекта и эталонных координат до и после согласования графа проективных преобразований, а именно

$$C(A_1, \dots, A_n, B_1, \dots, B_n) = \frac{1}{n} \sum_{j=1}^n d(A_j, B_j), \quad (5)$$

где  $A$  – координаты четырехугольников до и после согласования графа,  $B$  – эталонные координаты.

В эксперименте выбросы моделировались на некоторых кадрах видеопоследовательности (подробнее описано ниже), поэтому они известны заранее. Для оценки точности и полноты детек-



Рис. 3. График оценки качества при  $\sigma = 40$ .

тирования выбросов (Olson et al., 2008) используются метрики *precision* и *recall*

$$Precision = \frac{TP}{TP + FP}, \quad (6)$$

$$Recall = \frac{TP}{TP + FN}, \quad (7)$$

где  $TP$  – количество корректно определенных выбросов,  $FP$  – количество ложноопределенных выбросов среди не выбросов,  $FN$  – количество ложноопределенных не выбросов при выбросах.

## ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ

Для проведения экспериментов взята база данных MIDV-500, содержащая в себе разбитые на кадры видеопоследовательности с изображением документов, удостоверяющих личность, разных стран. Каждая видеопоследовательность разбита на 30 кадров. Использованы документы с кодом: 01–05, 07, 09–10, 12–14, 16–17, 19–24, 26–38, 40–50.

Для проведения экспериментов создана искусственная среда, для чего моделировался выбросовый шум на отдельно выбранных кадрах при помощи случайной величины по нормальному распределению  $N_2$ , которая добавлялась к каждой координате эталонного расположения объекта на кадре

$$q_i \leftarrow q_i + X_i : X_i \sim N_2(0, \sigma^2 I_2), \quad (8)$$

где  $\sigma$  – стандартное отклонение от нормального распределения,  $I_2$  – единичная матрица.

Составлена таблица для сравнения результатов по оценке качества работы алгоритма (табл. 1). Номер означает код страны в базе данных, на котором был проведен эксперимент. Качество с

меньшим значением считается лучшим. В таблице наглядно видно улучшение качества после согласования графа преобразований по сравнению с результатом до использования предложенного. Приведенные результаты были получены при эксперименте с значением  $\sigma = 40$ , выброс присутствовал на 5- и 15-м кадрах. Клипы содержат в себе изображение прямоугольного документа на фоне стола.

Значения  $precision(6)$  и  $recall(7)$  для задачи детектирования выбросов среди кадров анализированных видеопоследовательностей составляют 0.562 и 0.782 соответственно. Как видно из табл. 1, алгоритм имеет достаточно много ложных выбросов, однако эти кадры чаще всего являются соседними с выбросами. Это обусловлено тем, что при согласовании графа матриц проективных преобразований координаты соседних кадров корректируются с участием координат выбросовых кадров. Данная ситуация не ухудшает качество детектирования плоского объекта в целом и точность классификации выбросовых кадров достаточно велика.

Для наглядности результатов построен график оценки качества для всех экспериментов, представленный на рис. 3. Помимо этого на графике отмечено среднее значение качества до и после использования метода согласования графа преобразований.

Прямыми линиями с цифрами 1 и 2 изображены средние значения для всего эксперимента до и после согласования графа преобразований соответственно. Ломаными линиями отмечены значения качества работы алгоритма на каждом эксперименте. Как видно на графике, предлагаемый метод фильтрации дает положительный результат и улучшает качество детектирования объекта, так как среднее значение оценки уменьшилось.

Таблица 1. Результаты экспериментов при  $\sigma = 40$ 

Номер	Качество до согласования графа	Качество после согласования графа	Кадры, определенные как выброс
1	2.73415	1.37987	5, 15
2	3.39411	1.98901	5, 15
3	3.58267	2.39666	5, 15, 18
4	2.63987	1.78376	5, 6, 15
5	2.54558	2.27926	5, 14, 15, 25
7	0.848528	0.848528	—
9	3.48839	1.75645	5, 15
10	1.88562	2.78103	5, 15, 22, 26, 27, 28
12	2.16846	2.16846	—
13	1.03709	0.705505	5, 15
14	2.82843	2.98573	4, 5, 14, 15, 18
16	1.88562	2.51231	20
17	2.92271	2.1552	15
19	2.82843	2.75071	5, 7, 15, 16
20	1.88562	2.46925	15, 25
21	2.73415	1.12255	5, 15
22	1.88562	1.56072	5, 15, 28
23	2.16846	1.0047	5, 15
24	2.4513	1.93083	5, 15, 22
26	1.60278	1.11921	15
27	2.54558	3.02695	5, 15, 21, 22, 27
28	2.82843	2.42447	5, 6, 15, 16
29	2.07418	1.6642	4, 5, 15, 16
30	3.39411	2.71167	5, 15, 20
31	1.22565	2.91169	23, 26, 27
32	3.01699	1.98958	5, 15
33	2.73415	2.81777	5, 6, 11, 15
34	1.22565	1.42706	5, 15, 16, 17
35	3.48839	3.76994	5, 15, 24, 27, 29
36	3.48839	2.0424	5, 15
37	3.11127	2.30831	5, 6, 15
38	3.20555	1.80131	5, 15
40	3.01699	2.4152	15
41	3.39411	4.56123	5, 11
42	2.92271	2.90484	4, 5, 15, 16
43	3.11127	2.87427	5, 6, 7, 15
44	1.77102	1.48509	5, 15
45	2.4513	2.34363	5, 6, 7, 15
46	2.1343	1.76286	5, 15, 28
47	2.73415	3.77742	5, 15, 21, 24, 27
48	0.377124	1.28605	19, 25, 28
49	0.848528	1.61455	3, 4, 16
50	1.31993	1.01027	15

В случаях, где качество детектирования ухудшилось, алгоритм ошибочно выбирал выбросовый шум из-за нечетких кадров и некорректной работы построения матриц проективных преобразований. Также причиной неправильно выбранных кадров может быть плохое освещение, из-за чего ошибочно определялись особые точки.

### ЗАКЛЮЧЕНИЕ

В работе предложен метод фильтрации выбросов в процессе локализации прямоугольных плоских объектов при видеосъемке при помощи общей системы координат. Метод основывается на согласовании графа матриц проективных преобразований между кадрами, что позволяет выявить выбросовые кадры и откорректировать полученные координаты расположения объекта. Метод протестирован на реальных данных и показал хорошую точность классификации (78%) выбросовых кадров, но при этом часто соседние кадры также определяются как выбросовые, но данная ситуация не ухудшает результат детекции в целом. Для того, чтобы улучшить качество работы предлагаемого алгоритма, в дальнейших работах предлагаем усовершенствовать метод и использовать повторное согласование графа матриц проективных преобразований, при этом удаляя ребра из графа, соединяющие нулевую вершину с выбросовыми вершинами. Предполагается улучшение коррекции координат расположения объекта на кадрах.

Работа выполнена при частичной финансовой поддержке РФФИ (проекты 18-07-01387 и 18-29-26037)

### СПИСОК ЛИТЕРАТУРЫ

Савельев Б.И., Мамай И.Б., Николаев Д.П., Арлазаров В.Л., Булатов К.Б., Скорюкина Н.С. Метод согласования графа проективных преобразований для задачи панорамирования плоских объектов. *Тр. и-та системного анализа*. 2018. Т. 68. № S1. С. 124–133.

Хартсхорн Р. *Основы проективной геометрии*. М.: Мир, 1970. 160 с.

- Babu N., Soumya A. Character recognition in historical handwritten documents. *Proceedings of the 2019 IEEE International Conf. Communic. and Signal Processing*. 2019. P. 299–304.
- Fan B., Qingqun K., Tomasz T., Zhiheng W. Receptive fields selection for binary feature description. *IEEE Transactions on Image Processing*. 2014. V. 23. № 6. P. 2583–2595.
- Iglewicz B., Hoaglin D. How to Detect and Handle Outliers. *The ASQC Basic References in Quality Control: Statistical Techniques*. Eds F. Edward, P. Mykytka. 1993. 77 p.
- Kelson R.T., Andre M.S., Adelardo A.D. Optical Flow Using Color Information. *ACM New York*. NY, USA. 2008. P. 5–10.
- Lukoyanov A.S., Nikolaev D.P., Konovalenko I.A. Modification of YAPE keypoint detection algorithm for wide local contrast range image. *Information technologies and nanotechnology*. 2018. P. 1193–1204.
- Martin A.F., Robert C.B. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. of the ACM*. 1981. V. 24. P. 381–395.
- Moreno-García C.F., Elyan E., Jayne C. New trends on digitisation of complex engineering drawings. *Neural Computing and Applications*. 2019. V. 31 (6). P. 1695–1712.
- Newman P., Ho K. SLAM-loop closing with visually salient features. *IEEE Proc. of International Conference on Robotics and Automation*. 2005. P. 635–642.
- Novack G.D., Lim M.C. Retinal Detachment: Patient Perspective and Electronic Health Records. *American Journal of Ophthalmology*. 2019. V. 208. P. 64–67.
- Olson D. L. Delen D. Advanced Data Mining Techniques. *Springer, 1st edition*. 2008. 138 p.
- Skoryukina N., Nikolaev D., Arlazarov V. 2D art recognition in uncontrolled conditions using one-shot learning. *ICMV*. 2018. P. 1–8.
- Skoryukina N., Nikolaev D.P., Sheshkus A., Polevoy D. Real time rectangular document detection on mobile devices. *In Seventh International Conference on Machine Vision*. 2015. V. 9445. P. 1–6.
- Turcot P., Lowe D.G. Better matching with fewer features: The selection of useful features in large database recognition problems. *Computer Vision Workshops (ICCV Workshops)*. 2009. P. 2109–2116.

## Filtering erroneous frame-by-frame results for planar rectangular objects localization in a video using matching transform graph

E. V. Emelyanova<sup>a,b,✉</sup>, B. I. Saveliev<sup>b</sup>, and K. B. Bulatov<sup>b,c</sup>

<sup>a</sup> National Research Technological University “MISiS”  
119991, Moscow, Leninsky prospect, 4, Russia

<sup>b</sup> Smart Engines Service LLC  
117312 Moscow, pr. 60-letiya Oktyabrya, 9, Russia

<sup>c</sup> Federal Research Center “Informatics and Management” of the Russian Academy of Sciences  
127051 Moscow, B. Karetny per. 19, Russia

<sup>✉</sup>E-mail: e.emelianova@smartengines.ru

The detection of rectangular objects is an important part of document recognition and workflow automation. At the same time, outlier noise may appear during the detection process – falsely determined coordinates of the document location, negatively affecting the overall quality of object detection and further processing. The task of filtering such noise when detecting a flat rectangular object on frames of a video sequence by constructing a single coordinate system is posed. The proposed method is based on the coordination of the graph of estimated projective transformations of frames with templates and frames among themselves. In the experiments used, simulated outlier applied to some frames of the video sequence. To assess the quality of the algorithm, the results of object location before and after the module matching the graph with the true location of the object are compared. This assessment is based on the open MIDV-500 database of video filming of identity documents. This data set demonstrates a reduction in the cumulative error compared to the results before using graph matching in calculating projective transformations.

**Key words:** graph adjustment, projective transform, common coordinate system, video stream, object detection, cumulative error

## REFERENCES

- Babu N., Soumya A. Character recognition in historical handwritten documents. *Proceedings of the 2019 IEEE International Conference on Communication and Signal Processing. ICCSP 2019.* P. 299–304.
- Fan B., Qingqun K., Tomasz T., Zhiheng W. Receptive fields selection for binary feature description. *IEEE Transactions on Image Processing.* 2014. V. 23. № 6. P. 2583–2595.
- Hartshorn R. *Osnovy proektivnoy geometrii* [Fundamentals of projective geometry]. M.: Mir, 1970. 160 p. (in Russian).
- Iglewicz B., Hoaglin D. How to Detect and Handle Outliers. *The ASQC Basic References in Quality Control: Statistical Techniques.* Eds F. Edward, P. Mykytka. 1993. 77 p.
- Kelson R.T., Andre M.S., Adelardo A.D. Optical Flow Using Color Information. *ACM New York.* NY, USA. 2008. P. 5–10.
- Lukoyanov A.S., Nikolaev D.P., Konovalenko I.A. Modification of YAPE keypoint detection algorithm for wide local contrast range image. *Information technologies and nanotechnology.* 2018. P. 1193–1204.
- Martin A.F., Robert C.B. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. of the ACM.* 1981. V. 24. P. 381–395.
- Moreno-García C.F., Elyan E., Jayne C. New trends on digitisation of complex engineering drawings. *Neural Computing and Applications.* 2019. V. 31 (6). P. 1695–1712.
- Newman P., Ho K. SLAM-loop closing with visually salient features. *IEEE Proc. of International Conference on Robotics and Automation.* 2005. P. 635–642.
- Novack G.D., Lim M.C. – Retinal Detachment: Patient Perspective and Electronic Health Records. *American Journal of Ophthalmology.* 2019. 208. P. 64–67.
- Olson D.L. Delen D. Advanced Data Mining Techniques. *Springer, 1st edition.* 2008. 138 p.
- Savelyev B.I., Mamay B., Nikolaev D.P., Arlazarov V.L., Bulatov K.B., Skoryukina N.S. Metod soglasovaniy grafa proektivnykh preobrazovanii dlya zadachi panoramirovaniy obyektov. [A method of projective transformations graph adjustment for panorama stitching problem for images of planar objects]. *Proceedings ISA RAS.* 2018. V. 68. № S1. P. 124–133 (in Russian).
- Skoryukina N., Nikolaev D., Arlazarov V. 2D art recognition in uncontrolled conditions using one-shot learning. *ICMV.* 2018. P. 1–8.
- Skoryukina N., Nikolaev D.P., Sheshkus A., Polevoy D. Real time rectangular document detection on mobile devices. *In Seventh International Conference on Machine Vision.* 2015. V. 9445. P. 1–6.
- Turcot P., Lowe D.G. Better matching with fewer features: The selection of useful features in large database recognition problems. *Computer Vision Workshops (ICCV Workshops).* 2009. P. 2109–2116.