

## МЕТОДЫ СОЗДАНИЯ ФОНОВЫХ ИЗОБРАЖЕНИЙ И ИХ ИСПОЛЬЗОВАНИЕ ПРИ КОНВЕРТАЦИИ 2D-ВИДЕО В 3D-ФОРМАТ

Н. Н. Красильников<sup>а</sup>, доктор техн. наук, профессор

О. И. Красильникова<sup>а</sup>, канд. техн. наук, доцент

<sup>а</sup>Санкт-Петербургский государственный университет аэрокосмического приборостроения

**Постановка проблемы:** технология конвертации 2D-видео в 3D-формат включает в себя две трудоемкие операции, выполнение которых, как правило, требует больших затрат ручного труда. К этим операциям относятся заполнение областей открытия, возникающих в результате смещения объектов сцены, обусловленного параллаксом, и выделение силуэтов объектов сцены путем семантической сегментации изображений конвертируемой последовательности кадров в процессе создания карты глубины. **Цель исследования:** разработка методов получения фоновых изображений конвертируемых последовательностей кадров и на их основе технологий заполнения областей открытия и выделения силуэтов изображений движущихся объектов. **Результаты:** разработаны методы создания фоновых изображений: метод предварительной съемки фонового изображения той же камерой, которая используется для съемки клипа; метод поиска фонового изображения в отснятом клипе и метод монтажа фонового изображения из фрагментов клипа для различных направлений движения объекта, — а также метод заполнения областей открытия, возникающих при конвертации 2D-изображений клипа в 3D-формат, с использованием для этой цели фонового изображения и метод выделения силуэтов изображений движущихся объектов, необходимых для создания карты глубины, из последовательности кадров клипа путем сопоставления изображений последовательности с фоновым изображением. **Практическая значимость:** разработанные методы позволяют при конвертации 2D-видео в 3D-формат сокращать трудоемкость процесса в десятки и более раз.

**Ключевые слова** — стереоскопическое изображение, фоновое изображение, область открытия, карта глубины.

### Введение

В настоящее время одной из причин, тормозящих развитие и внедрение 3D-кино и телевидения, является недостаточное количество качественного 3D-контента, что объясняется трудностями его создания [1]. Непосредственная съемка 3D-фильма, по самым грубым и оптимистичным подсчетам, увеличивает изначальный бюджет 2D-фильма на 30–50 % и более. Поэтому наряду с методом непосредственной съемки фильма в стереоскопическом (3D) формате был разработан метод конвертации 2D-фильмов в 3D-формат. В настоящее время используются оба метода получения 3D-контента, но и там и там возникают проблемы. В первом случае наиболее сложными являются проблемы выбора параметров съемки, фокусного расстояния объектива, расстояния между оптическими осями объективов стереокамеры и т. д., а также редактирования уже отснятого материала в целях коррекции допущенных ошибок при выборе исходных параметров съемки. Во втором случае проблемы обусловлены как необходимостью большого объема ручного труда при выполнении конвертации, так и отсутствием полной информации об изображении конвертируемой сцены, полученном из одной точки съемки.

В настоящее время около 2/3 стереоскопических фильмов производится путем конвертации фильмов, снятых в формате 2D, в формат 3D [2, 3].

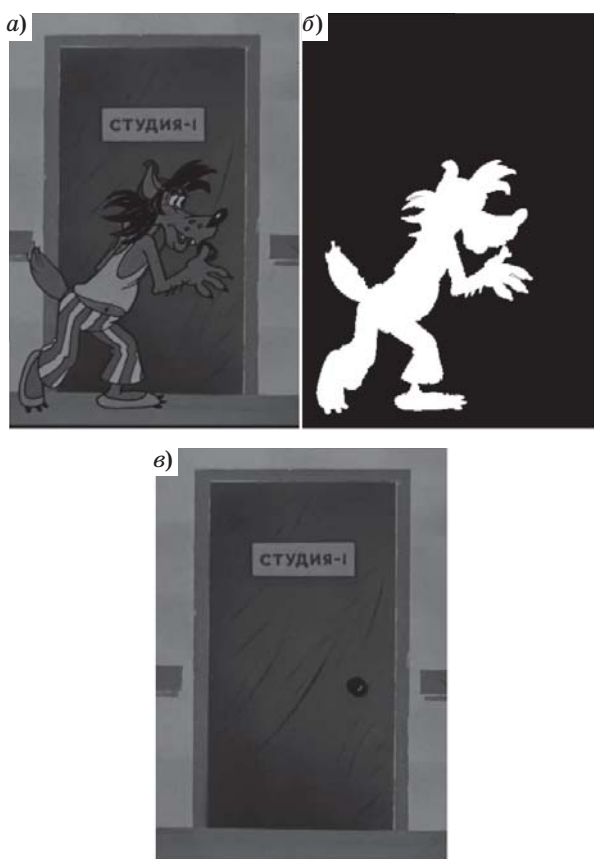
Одной из наиболее трудоемких операций конвертации последовательности 2D-изображений клипа фильма в 3D-формат является выполняемая вручную операция заполнения областей открытия, возникающих в результате обусловленного параллаксом смещения объектов сцены. Другой не менее трудоемкой операцией является сегментация, в результате которой выделяются изображения объектов на 2D-изображениях из фоновой части. Трудоемкость обусловлена тем, что, как правило, изображения объектов имеют сложную форму, которая к тому же изменяется при переходе от одного кадра к другому кадру видеопоследовательности. В связи с этим сегментация может быть только семантической, а следовательно, должна выполняться также вручную.

### Проблема заполнения областей открытия

Рассмотрим проблему заполнения областей открытия вследствие смещения объектов, вызванного экранном параллаксом. При конвертации, в результате смещения изображений объектов вправо и влево относительно своего первоначального положения на исходном 2D-изображении, на изображениях стереопары образуются области, которые должны быть заполнены изображением фона, который на исходном изображении заслонен объектом, а поскольку такое изображение отсутствует, то образуются так называемые



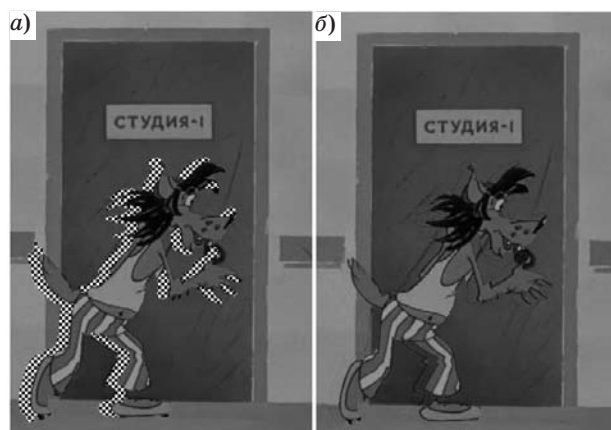
■ *Рис. 1.* Пояснение к возникновению областей открытия



■ *Рис. 2.* Метод заполнения областей открытия фоновым изображением: *а* — исходное изображение, включающее в себя объект и фоновую часть; *б* — силуэт объекта; *в* — фоновое изображение

области открытия, и их необходимо заполнить. Одна из таких областей показана на рис. 1.

Часто эта проблема решается путем повторения пикселя на границе области открытия



■ *Рис. 3.* Результаты конвертации 2D-изображения в 3D-формат для левого компонента стереоскопического изображения, на котором области открытия видны (*а*); заполнены фрагментами фонового изображения (*б*)

(методом интерполяции нулевого порядка). Для этой цели предлагаются также метод линейной интерполяции, метод текстурного заполнения и ряд других [4–7]. К сожалению, ни один из них не позволяет получать удовлетворительные результаты.

Для полноценного решения проблемы заполнения областей открытия необходимо располагать фоновым изображением, которое не содержало бы движущиеся объекты [3, 8–11]. В этом случае области открытия могут быть просто заполнены соответствующими фрагментами фонового изображения. Таким образом, проблема заполнения областей открытия может быть сведена к более простой задаче — получению фонового изображения. Попутно отметим, что для этого не обязательно иметь все фоновое изображение, а достаточно располагать только областями, совпадающими с областями открытия, и проблема заполнения будет решена.

Кроме того, наличие фонового изображения упрощает создание карты глубины путем сравнения каждого из кадров конвертируемой видеопоследовательности с фоновым изображением и получения таким образом силуэтов изображений объектов, которые впоследствии, после соответствующей обработки, будут заполнены пикселями необходимой интенсивности. Проведенные нами исследования на видеопоследовательностях, взятых из реальных фильмов, показали, что в значительной части случаев фоновые изображения, на которых отсутствуют изображения объектов, могут быть сравнительно несложным способом получены из самой видеопоследовательности. Рис. 2, *а–в* и 3, *а, б* поясняют описываемый метод заполнения областей открытия.

## Создание фонового изображения

Самым простым способом получения фонового изображения является предварительная съемка сцены камерой в отсутствие движущихся объектов. При этом необходимо, чтобы:

- съемка фонового изображения производилась той же камерой, что и съемка всего снимаемого сюжета (клипа), для которого снимается фоновое изображение;

- положение камеры было зафиксировано и не изменялось бы в процессе съемки клипа;

- при съемке фонового изображения и последующего клипа сохранялись бы неизменными все настройки камеры и режим освещения. Выполнение перечисленных условий может оказаться не таким простым, как представляется, особенно при съемке протяженных по времени клипов. Даже при съемках в студии трудно обеспечить неизменным режим освещения, поскольку перемещающиеся объекты, отбрасывая тени и переотражая свет от источников освещения, нарушают его.

Часто возможность произвести предварительную съемку фоновых изображений отсутствует, например, когда необходимо конвертировать уже отснятый фильм. В этом случае фоновые изображения можно создать путем сборки из фрагментов уже имеющихся последовательностей, на которых фон не заслонен движущимися объектами (рис. 4). На рис. 4, *а* и *б* приведены два кадра из мультфильма «Ну, погоди!», содержащие движущийся объект (волка), но эти кадры выбраны из видеопоследовательности таким образом, что движущийся объект на них в обоих случаях заслоняет различные части фона. Благодаря этому путем монтажа было составлено фоновое изображение (рис. 4, *в*). При выборе кадров видеопоследовательности, из которых в дальнейшем получается фоновое изображение, необходимо следить за тем, чтобы они были сняты при неизменном положении камеры. Это

выполняется путем их сравнения. В некоторых случаях возникает необходимость небольшого редактирования получаемого фонового изображения с тем, чтобы сделать незаметными места сшивки.

Теперь рассмотрим создание фонового изображения, когда в последовательности кадров клипа отсутствует хотя бы один кадр, который не содержит движущийся объект, но объект движется в направлении на камеру или от камеры, закрывая собой часть фона. Для реализации этого случая необходимо выполнить следующее:

- если объект перемещается на камеру, нужно, начиная со второго кадра от начала последовательности, изготовить ряд силуэтов движущегося объекта (их необходимое количество определяется опытным путем);

- определить для изображений этого объекта значения экранных параллаксов;

- путем морфологической обработки вида «эрозия» изготовленных силуэтов движущегося объекта удалить с каждого из них количество слоев пикселей, равное половине экранного параллакса;

- путем сравнения слоев посредством наложения подобрать ближайший к первому кадру кадр последовательности, в котором изображение движущегося объекта первого кадра полностью помещается внутри силуэта, предварительно подвергнутого эрозии. Начиная с этого ближайшего кадра и до конца последовательности изображение первого кадра последовательности можно использовать как фоновое.

В том случае если объект при движении удаляется от камеры, описанные выше действия необходимо выполнять, начиная с предпоследнего кадра последовательности, т. е. с кадра, на котором изображение движущегося объекта имеет небольшие размеры (рис. 5, *а-в* и 6, *а, б*).

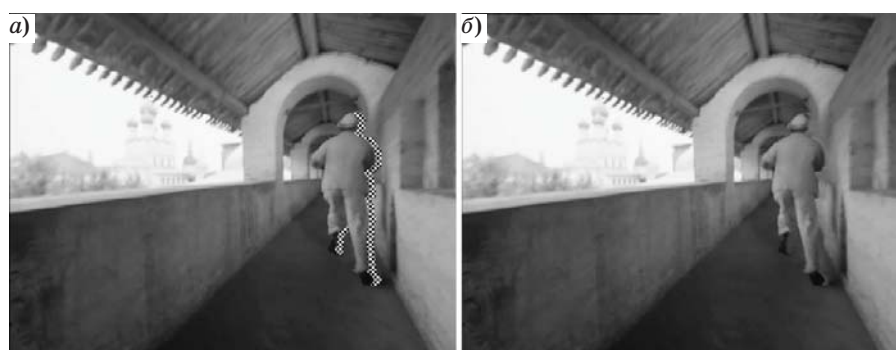
Для кадров, заключенных между кадром, который используется как фоновое изображение, и кадром, с которого фоновое изображение начинает использоваться для заполнения областей



■ **Рис. 4.** Создание фонового изображения из двух кадров: *а, б* — исходные изображения, включающие в себя объект (волка в разных местах коридора и фоновую часть); *в* — фоновое изображение, полученное из первых двух путем монтажа



■ **Рис. 5.** Создание фонового изображения из двух кадров: *а, б* — исходное изображение, на котором движущийся объект (человек) находится соответственно относительно близко к камере и далеко от камеры; *в* — силуэт этого движущегося объекта



■ **Рис. 6.** Результат конвертации исходного изображения в 3D-формат: на левом компоненте стереоскопического изображения области открытия видны (*а*) и заполнены фрагментами фонового изображения (*б*)

открытия, ничего другого не остается, как заполнить эти области одним из традиционных методов [3–7].

### Адаптация фонового изображения к изменяющимся условиям съемки

Как уже было отмечено, фоновый компонент изображения в различных кадрах конвертируемой последовательности может быть подвержен небольшим изменениям, обусловленным незначительными варьированиями как освещения снимаемой сцены, так и положения снимающей камеры. Это особенно проявляется при съемке клипов, имеющих значительную протяженность во времени. Величины смещений фонового изображения в течение нескольких секунд, как показали измерения на реальных фильмах, не превышают одного пикселя. Изменения яркости пикселей фонового компонента изображения кадра при наличии в кадре движущихся объектов даже за это время могут быть вполне ощутимыми, особенно если движущиеся объекты отбрасывают тени.

Для компенсации рассогласования фоновых изображений, созданных одним из описанных выше методов, с фоновыми компонентами изображений, в которых производится заполнение областей открытия, необходимо производить их адаптацию к тем изменениям, которые имели место при съемке конвертируемых изображений кадров.

Вначале необходимо для каждого кадра конвертируемой последовательности изготовить индивидуальные фоновые изображения в предположении, что рассогласования имеют место только в виде смещений очередного конвертируемого кадра последовательности относительно исходного фонового изображения. При этом процедура компенсации будет состоять в смещении исходного фонового изображения до полного совпадения по столбцам и по строкам с фоновым компонентом изображения, в котором производится заполнение областей открытия.

После этого необходимо компенсировать рассогласование по яркости каждого изготовленного описанным методом индивидуального фонового изображения с фоновым компонентом изображения, в котором производится заполнение обла-

стей открытия. Операция эта выполняется с самого начала конвертируемой последовательности кадров в соответствии с правилом

$$L_{\Phi}(k, n, i) = L_c(k, n, i),$$

если  $|L_c(k, n, i) - L_{\Phi}(k, n, i)| < P,$

где  $L_{\Phi}(k, n, i)$  — индивидуальное фоновое изображение  $i$ -го кадра последовательности;  $L_c(k, n, i)$  — изображение  $i$ -го кадра конвертируемой последовательности;  $k$  — номер строки;  $n$  — номер столбца;  $i$  — номер кадра;  $P$  — порог, который подбирается экспериментально [12].

### Создание силуэта изображения движущегося объекта

Как уже было отмечено, наиболее сложной операцией при конвертации 2D-видео в 3D-формат является выделение силуэтов движущихся объектов из изображений кадров конвертируемой последовательности. Обусловлено это тем, что вследствие сложности контуров движущихся объектов эту операцию приходится выполнять вручную. Наличие фоновых изображений существенно упрощает проблему, поскольку операция выделения силуэтов движущихся объектов может быть в значительной степени автоматизирована путем сравнения изображения кадра, который содержит движущийся объект, с соответствующим фоновым изображением, и формирования на основе этого заготовки для изготовления силуэта (рис. 7, а-в). Выполняется это в соответствии с правилом

$$S(k, n, i) = \begin{cases} 255, & \text{если } |L_c(k, n, i) - L_{\Phi}(k, n, i)| \geq P, \\ 0, & \text{если } |L_c(k, n, i) - L_{\Phi}(k, n, i)| < P \end{cases},$$

где  $S(k, n, i)$  — заготовка для изготовления силуэта изображения движущегося объекта.

Из рисунка видно, что вследствие флуктуационного шума, присутствие которого неизбежно как на изображении кадра, содержащего движущийся объект, так и на фоновом изображении, возникают характерные погрешности изображения заготовки силуэта, которые проявляются в виде незаполненных областей, нарушения гладкости контура, а также в виде отдельных выбросов шума на полях, которые должны оставаться незаполненными. Поэтому заготовка нуждается в ручном редактировании, которое заключается в заполнении незаполненных областей силуэта, сглаживании контуров и устранении светлых точек и штрихов в остальной части поля. Однако эта операция несравненно более простая, чем рисование силуэта движущегося объекта. Это особенно важно при конвертации видеопоследовательностей, протяженность которых часто превышает сотни кадров. Естественно, что перед тем как приступить к выполнению операции выделения силуэтов движущихся объектов, из изображений последовательности кадров, которые их содержат, следует произвести адаптацию фоновых изображений к изменяющимся условиям съемки в соответствии с описанным выше методом.

### Заключение

В заключение отметим, что длительность клипов (отдельных сюжетов), из которых состоит видеопоследовательность реальных фильмов, обычно заключена в интервале от 3–5 с и до десятков секунд, следовательно, в пересчете на количество кадров, формирующих такую последовательность при частоте кадров 25 кадров/с, это составит от 75 и до нескольких сотен и даже тысяч кадров. Из этого следует, что метод заполнения областей открытия путем применения предварительно созданных для этой цели фоновых изображений позволяет существенно сократить объем



■ **Рис. 7.** Создание заготовки силуэта движущегося объекта: а — исходное изображение, включающее в себя движущийся объект (волка) и фоновую часть; б — фоновое изображение; в — изображение заготовки силуэта

работы по сравнению с методом заполнения этих областей ручным способом, поскольку метод интерполяции и другие применяемые в настоящее время методы не приводят к хорошим результатам.

Формирование фонового изображения позволяет также весьма существенно упростить возникающую при создании карты глубины проблему сегментации изображений движущихся объектов путем выделения их силуэтов.

## Литература

1. Ватолин Д. С., Боков А. А., Федоров А. А. Тенденции изменения технического качества стереокино — 5 лет после «Аватара» // Мир техники кино. 2015. № 3(37). С. 17–28.
2. Real 3D or Fake 3D. <http://realorfake3d.com/> (дата обращения: 15.05.2017).
3. Ерофеев М. В., Ватолин Д. С. Многослойное решение проблемы полупрозрачных границ при построении стереоскопических изображений // Intern. Journal of Open Information Technologies. 2016. N 8(4). <http://istina.msu.ru/media/publications/article/854/979/27852059/310-900-1-PB.pdf>. (дата обращения: 15.05.2017).
4. Acharya T., Tsai P. S. Computational Foundations of Image Interpolation Algorithms // ACM Ubiquity. 2007. N 8. <http://ubiquity.acm.org/article.cfm?id=1317488> (дата обращения: 15.05.2017).
5. Красильников Н. Н., Красильникова О. И. Методы конвертации 2D-изображений и видео в стереоскопический формат // Информационно-управляющие системы. 2015. № 5. С. 18–25. doi:10.15217/issn1684-8853.2015.5.18
6. Красильников Н. Н., Красильникова О. И. Исследование восприятия 3D-объектов зрительной системой человека для разработки методов конвертации 2D-изображений в стереоскопические изображения // Оптический журнал. 2015. № 10(82). С. 1–8.
7. Newson A., Almansa A., Fradet M., Gousseau Y., Pérez P. Video Inpainting of Complex Scenes // SIAM Journal on Imaging Sciences. 2014. N 4(7). P. 1993–2019. doi:10.1137/140954933
8. Ерофеев М. В., Ватолин Д. С. Матирование видео-последовательностей с использованием восстановленного фона // Цифровая обработка сигналов. 2016. № 3. С. 19–25.
9. Wang J., Cohen M. F. Image and Video Matting: a Survey // Foundations and Trends® in Computer Graphics and Vision. 2007. N 2(3). P. 97–175. doi:10.1561/06000000019
10. Rhemann C., Rother C., Wang J., Gelautz M., Kohli P., Rott P. A Perceptually Motivated Online Benchmark for Image Matting // Computer Vision and Pattern Recognition: Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR 2009), Miami, Florida, USA. 2009. P. 1826–1833. doi:10.1109/CVPR.2009.5206503
11. Lee S.-Y., Yoon J.-C., Lee I.-K. Temporally Coherent Video Matting // Graphical Models. 2010. N 3(72). P. 25–33. doi:10.1016/j.gmod.2010.03.001
12. Красильников Н. Н. Цифровая обработка 2D- и 3D-изображений. — СПб.: БХВ-Петербург, 2011. — 608 с.

UDC 612.843.721

doi:10.15217/issn1684-8853.2017.3.11

### Creating Background Images and Their use in Converting 2D video into 3D format

Krasilnikov N. N.<sup>a</sup>, Dr. Sc., Tech., Professor, NNKrasilnikov@yandex.ru

Krasilnikova O. I.<sup>a</sup>, PhD, Tech., Associate Professor, OIKrasilnikova@yandex.ru

<sup>a</sup>Saint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaja St., 190000, Saint-Petersburg, Russian Federation

**Introduction:** The technology of converting 2D video into 3D format includes two time-consuming operations which usually require a lot of manual labor. The first operation is filling up the opening areas which result from the scene object displacement caused by the parallax. The second one is the selection of scene object silhouettes by semantic segmentation of the frame sequence during the creation of the depth map. **Purpose:** We develop methods for obtaining background images of the frame sequences to be converted. These methods can be the base for a technology of filling up the opening areas and for a technology of allocating moving object silhouettes. **Results:** We have developed methods for creating background images, namely: a method for preliminary shooting of the background image by the same camera which is used for shooting the clip, a method of searching for the background image in the captured clip, and a method for assembling the background image from fragments of the clip for various directions of the object movement. Also, methods have been developed for filling up the opening areas which arise when converting 2D images of a clip into 3D format by using the background image for this purpose, and for allocating moving object image silhouettes (which are necessary for creating a depth map) from the clip frame sequence by comparing the images on the sequence with the background image. **Practical relevance:** The developed methods allow you to reduce the complexity of converting 2D videos into 3D format tenfold or more.

**Keywords** — Stereoscopic Image, Background Image, Opening Areas, Depth Map.

## References

1. Vatolin D. S., Bokov A. A., Fedorov A. A. Trends in the Technical Quality of Stereoscopic — 5 Years after “Avatar”. *Mir tekhniki kino*, 2015, no. 3(37), pp. 17–28 (In Russian).
2. *Real 3D or Fake 3D*. Available at: <http://realorfake3d.com/> (accessed 15 May 2017).
3. Erofeev M. V., Vatolin D. S. Multilayer Solution of the Problem of Translucent Boundaries in the Construction of Stereoscopic Images. *Intern. Journal of Open Information Technologies*, 2016, vol. 4, no. 8. Available at: <http://istina.msu.ru/media/publications/article/854/979/27852059/310-900-1-PB.pdf> (accessed 15 May 2017).
4. Acharya T., Tsai P. S. Computational Foundations of Image Interpolation Algorithms. *ACM Ubiquity*, 2007, no. 8. Available at: <http://ubiquity.acm.org/article.cfm?id=1317488> (accessed 15 May 2017).
5. Krasilnikov N. N., Krasilnikova O. I. Methods of Converting 2D Images and Videos into Stereoscopic Format. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2015, no. 5, pp. 18–25 (In Russian). doi:10.15217/issn1684-8853.2015.5.18
6. Krasilnikov N. N., Krasilnikova O. I. Study of the Perception of 3D Objects by the Human Visual System in Order to Develop Methods for Converting 2D Images into Stereoscopic Images. *Opticheskiy zhurnal*, 2015, no. 10(82), pp. 1–8 (In Russian).
7. Newson A., Almansa A., Fradet M., Gousseau Y., Pérez P. Video Inpainting of Complex Scenes. *SIAM Journal on Imaging Sciences*, 2014, no. 4(7), pp. 1993–2019. doi:10.1137/140954933
8. Erofeev M. V., Vatolin D. S. Matting Video Sequences using a Restored Background. *Tsifrovaia obrabotka signalov*, 2016, no. 3, pp. 19–25 (In Russian).
9. Wang J., Cohen M. F. Image and Video Matting: a Survey. *Foundations and Trends® in Computer Graphics and Vision*, 2008, no. 2(3), pp. 97–175. doi:10.1561/0600000019
10. Rhemann C., Rother C., Wang J., Gelautz M., Kohli P., Rott P. A Perceptually Motivated Online Benchmark for Image Matting. *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition “Computer Vision and Pattern Recognition” (CVPR 2009)*, Miami, Florida, USA, 2009, pp. 1826–1833. doi:10.1109/CVPR.2009.5206503
11. Lee S.-Y., Yoon J.-C., Lee I.-K. Temporally Coherent Video Matting. *Graphical Models*, 2010, no. 3(72), pp. 25–33. doi:10.1016/j.gmod.2010.03.001
12. Krasilnikov N. N. *Tsifrovaia obrabotka 2D- i 3D-izobrazhenii* [Digital Processing of 2D- and 3D-Images]. Saint-Petersburgs, BKhV-Peterburg Publ., 2011. 608 p. (In Russian).

## ПАМЯТКА ДЛЯ АВТОРОВ

*Поступающие в редакцию статьи проходят обязательное рецензирование.*

При наличии положительной рецензии статья рассматривается редакционной коллегией. Принятая в печать статья направляется автору для согласования редакторских правок. После согласования автор представляет в редакцию окончательный вариант текста статьи.

Процедуры согласования текста статьи могут осуществляться как непосредственно в редакции, так и по e-mail ([ius.spb@gmail.com](mailto:ius.spb@gmail.com)).

При отклонении статьи редакция представляет автору мотивированное заключение и рецензию, при необходимости доработать статью — рецензию. Рукописи не возвращаются.

*Редакция журнала напоминает, что ответственность за достоверность и точность рекламных материалов несут рекламодатели.*