

УДК 778.534.1

Звездаков С.В., Боков А.А., Ватолин Д.С.

Московский государственный университет
им. М.В. Ломоносова

АВТОМАТИЧЕСКАЯ МНОГОКЛАССОВАЯ КЛАССИФИКАЦИЯ ПЛАНОВ СТЕРЕОВИДЕО ПО МЕТОДУ ПРОИЗВОДСТВА

В данной работе предложен алгоритм, способный определять метод производства стереофильмов. Проанализировано 105 полнометражных стереоскопических фильмов, среди которых найдено 456 планов с конвертацией в 65 снятых фильмах и 86 планов со съёмкой в 40 сконвертированных фильмах.

Ключевые слова: стереоскопическое видео, контроль качества, конвертация, стереоскопическая съёмка, компьютерная графика, классификация планов.

Zvezdakov S.V., Bokov A.A., Vatolin D.S.

AUTOMATIC MULTICLASS S3D SHOT CLASSIFICATION BY PRODUCTION METHOD

In this article we propose an algorithm for stereoscopic video classification by the production method. We demonstrate the utility of the proposed method by applying it to 105 full-length 3D movies. In total, 456 converted scenes were found in 65 natively captured S3D movies and 65 captured scenes were found in 40 movies that used 2D-to-3D conversion as the main production method.

Key words: stereoscopic video, quality assessment, conversion, stereoscopic shooting, computer graphics, scene classification.

ВВЕДЕНИЕ

Несмотря на увеличение количества залов для демонстрации стереофильмов в кинотеатрах и выход большого количества фильмов в формате 3D, зрители всё реже ходят на стереофильмы. Согласно опросам, одной из основных причин этого явления является головная боль и неприятные ощущения зрителей после просмотра стереоскопических фильмов [2]. Одним из основных факторов, отрицательно влияющих на самочувствие зрителя при просмотре стереофильма, является его низкое техническое качество. Поэтому необходима оценка качества производимых стереоскопических фильмов.

Существует три основных метода создания стереоконтента: стереоскопическая съёмка, конвертация из двухмерного видео в стереоформат и создание средствами компьютерной графики. Для каждого метода производства характерны свои виды дефектов. Например, для съёмки это: всевозможные геометрические несоответствия, цветовые несоответствия, расхождения резкости и времени между ракурсами. Для конвертации основными дефектами являются: несоответствия резкости границ объектов в правом и левом ракурсах, эффект кулисности, несоответствия реальных и прорисованных границ объектов на карте диспаратности. Так как для разных методов производства стереофильмов характерны разные виды дефектов, которые требуется анализировать различными методами, то необходим алгоритм, способный по входному стереовидео определять метод его производства. В данной работе развивается идея алгоритма, предложенного в [1].

Анализ существующих методов. В 2015 году был предложен метод классификации стереовидео на конвертацию и стереоскопическую съёмку [4]. Авторы предполагают, что в сконвертированном стереовидео в местах заполнения образовавшихся областей открытия появится расхождение по резкости, вызванное некачественным заполнением. Существенные недостатки данного метода это ручной выбор кадров и объектов переднего плана для анализа областей открытия и низкая скорость работы (на обработку одного FullHD кадра тратится 55 секунд). Перечисленные недостатки делают этот метод неприменимым для анализа полнометражных стереофильмов.

1. ПРЕДЛАГАЕМЫЙ МЕТОД

В 2015 году авторами данной статьи разработан быстрый и полностью автоматический метод [1]. В нём для определения метода производства использовались следующие признаки:

- оценки расхождений между ракурсами VQMT3D [5];
- оценка размытия границ объектов на карте диспаратности;
- оценка трапецеидального искажения;
- оценка гладкости гистограммы диспаратности.

Затем по вычисленным признакам с помощью искусственных нейронных сетей производилась классификация метода производства на три описанных класса.

В данной работе предлагается модификация используемых признаков и классификатора для повышения точности классификации.

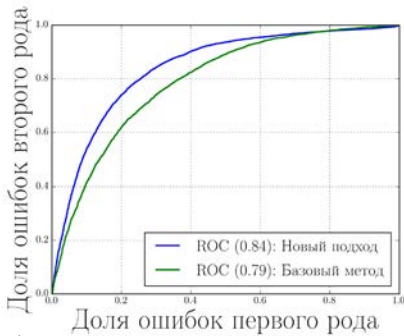
1.1. МОДИФИКАЦИЯ ОЦЕНКИ РАЗМЫТИЯ ГРАНИЦ ОБЪЕКТОВ НА КАРТЕ ДИСПАРАТНОСТИ

В базовом методе для преобразования накопленной по всему плану гистограммы модулей градиентов карты диспаратности использовалась часть гистограммы со значениями, большими определённого фиксированного порога. Из-за этого происходила потеря полезной для классификации информации. В данной работе для выделения наиболее информативной части гистограммы был применён следующий метод. Пусть значение карты диспаратности изменяется от 0 до 255. Гистограмма разбивается на корзины со значениями модуля градиента равными: 0, 1, 3, 5, 10, 15, 20, 30, 40, 50, 70, 100, 150, 200, 300, 400. При разбиении используется линейная интерполяция между корзинами, т. е. если значение модуля градиента равно 25, то его значение поровну попадёт в корзины 20 и 30. Далее рассматриваются различные комбинации корзин для решения задачи классификации стереофильмов только по этому признаку. Также рассматриваются варианты с нормированием гистограмм (при которых сумма всех значений гистограммы равна 1). Для решения задачи классификации использовалась логистическая регрессия, т. е. минимизировалась следующая функция:

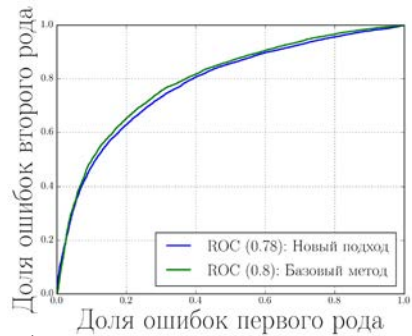
$$\min_{w, c} \left(\frac{1}{2} w^T w + \sum_{i=1}^n \log \left(\exp(-y_i (X_i^T w + c)) + 1 \right) \right),$$

где X_i^T — сектор-признак плана, y_i — сетка плана, а w и c — искомые параметры классификатора.

Результаты сравнения усовершенствованного подхода с базовым представлены на рис. 1, из которого видно, что пред-



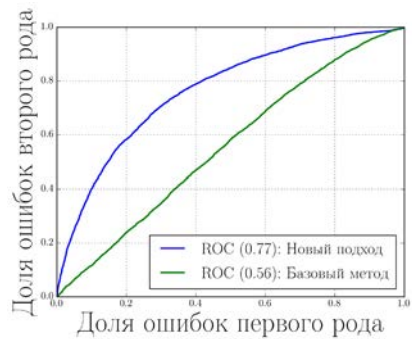
а)



б)

Рис. 1. Сравнение базового подхода с модифицированной версией алгоритма оценки размытия границ объектов:

а — конвертация/съёмка; б — графика/конвертация; в — графика/съёмка



в)

ложенный метод даёт лучшие результаты. По результатам сравнения в качестве признака для классификации была выбрана нормализованная гистограмма, из которой выбирается часть с корзинами: 0, 1, 5, 10, 15, 20, 30, 40, 50, 70, 100.

1.2. МОДИФИКАЦИЯ ОЦЕНКИ ТРАПЕЦЕИДАЛЬНОГО ИСКАЖЕНИЯ

Для преобразования накопленной гистограммы соответствия теоретического и реального вертикального сдвига в базовом методе использовалась весовая модификация метода RANSAC [3].

В предложенной модификации перед тем, как оценивать гистограммы линейной функцией, производится замена максимального значения в гистограмме вторым по размеру значением этой же гистограммы. Это необходимо для того, чтобы снизить вес корзины гистограммы, соответствующей нулевому значению.

Для построения приближающей функции используются два подхода: весовой и пространственный. Весовой подход предназ-

начен для построения модели, учитывающей веса корзин. В этом подходе для каждой корзины учитывается её удалённость d от модели с весом, равным

$$w(d) = \exp\left(\left(\frac{d}{0,375}\right)^8\right).$$

В результате этот подход строит модель, удовлетворяющую:

$$\min_M \left(\sum_{h_{i,j} \in H} w(d(i,j), M) \cdot h_{i,j} \right),$$

где M — искомая модель; $h_{i,j}$ — ячейка гистограммы, находящаяся на позиции (i, j) ; $d((i, j), M)$ — расстояние от точки (i, j) до модели M .

Пространственный подход рассматривает корзины гистограммы без учёта их веса. Гистограмма бинаризуется по порогу, равному 5% от максимального значения гистограммы. Точка не является выбросом, если она находится на расстоянии не более 0,25 пикселей от модели.

Этот подход предназначен для выбора модели, в которой находится больше всего непустых корзин. Результатом является модель, удовлетворяющая:

$$\max_M \left(\sum_{h_{i,j} \in H} I(d(i,j), M) \leq 0,25 \cdot I(h_{i,j} \geq 0,05 \cdot h_{\max}) \right),$$

где M — искомая модель; $h_{i,j}$ — ячейка гистограммы, находящаяся на позиции (i, j) ; $d((i, j), M)$ — расстояние от точки (i, j) до модели M ; I — функция-индикатор; h_{\max} — максимальное значение гистограммы.

Чтобы выбрать из полученных моделей лучшую, гистограмма линейно интерполируется с шагом 0,05 пикселя. Затем она бинаризуется по 5-процентному порогу. Из полученных моделей выбирается та, которая покрывает больше непустых корзин ин-

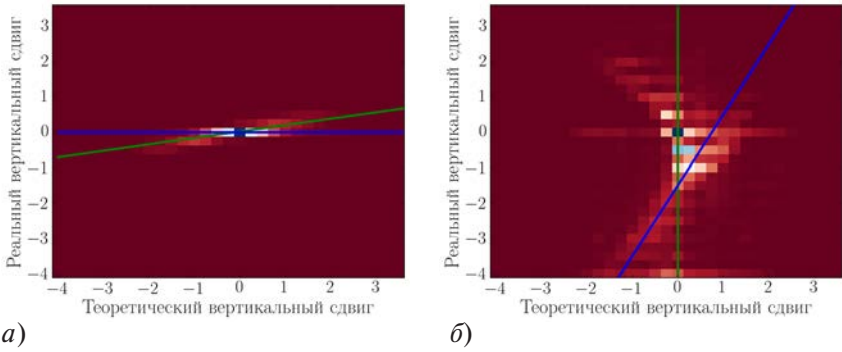


Рис. 2. Сравнение весового и пространственного подхода при построении линейной функции по гистограмме соответствия теоретического и реального вертикального сдвига (планы из фильма «Аватар»): *а* — план, на котором весовой подход показал лучшие результаты; *б* — план, на котором пространственный подход показал лучшие результаты

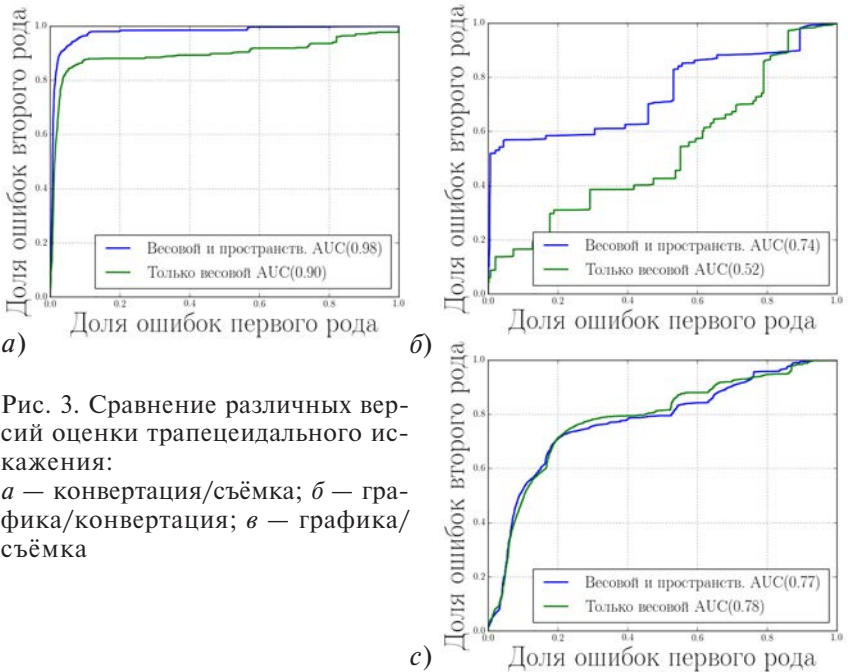


Рис. 3. Сравнение различных версий оценки трапецидального искажения: *а* — конвертация/съёмка; *б* — графика/конвертация; *в* — графика/съёмка

терполированной гистограммы на расстоянии 0,375 пикселя от модели. Сравнение подходов изображено на рис. 2: зелёная линия соответствует весовому подходу, синяя — пространственному. На рис. 2, *а* весовой подход показывает лучшие результаты, а на рис. 2, *б* — пространственный подход показывает лучшие результаты.

Сравнение модифицированного подхода с базовым представлено на рис. 3. Классификация производилась на каждую пару классов с использованием логистической регрессии.

1.3. ВЫБОРКА

Выборка состояла из 16 фильмов, размеченных полуавтоматически. Для увеличения корректности результатов тестирования выборка была разбита на три части: обучающая, тестируемая и валидационная. Обучение производилось на обучающей выборке с оценкой качества обучения на тестируемой выборке. Для оценки качества обученного классификатора использовалась валидационная выборка. Для увеличения адекватности результатов фильмы располагались либо в обучающей и тестируемой частях, либо в валидационной части.

2. ЭКСПЕРИМЕНТАЛЬНАЯ ОЦЕНКА

2.1. НОРМИРОВКА ДАННЫХ

В базовом методе лучшие результаты показали подходы с использованием искусственных нейронных сетей прямого распространения, поэтому в предложенной модификации в качестве классификатора используются именно они. Для увеличения скорости обучения и точности классификации искусственных нейронных сетей данные были нормированы. Опробованы следующие нормировки:

- нормализация стандартным отклонением,
- минимаксная нормализация.

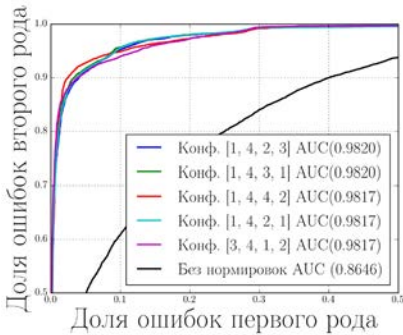
Для каждой оценки подбиралась своя нормировка, не зависящая от остальных оценок. Рассмотрены два варианта: нормировка по всем фильмам в обучающей выборке, нормировка по каждому фильму отдельно. Конфигурация нормировок для каждой оценки обозначается в виде: [оценки VQMT3D; оценка размытия границ объектов на карте диспаратности; оценка трапецеидального искажения; оценка гладкости гистограммы диспаратности]. Обозначения нормировок:

1. Нормализация стандартным отклонением по всем фильмам.
2. Нормализация стандартным отклонением по каждому фильму отдельно.
3. Минимаксная нормализация по всем фильмам.
4. Минимаксная нормализация по каждому фильму отдельно.

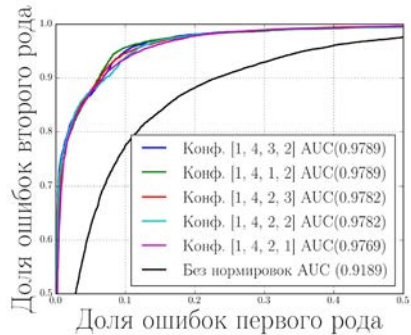
Для поиска наилучшего сочетания нормировок протестированы различные комбинации нормировок для каждого признака. Для классификации использовалась искусственная нейронная сеть с одним скрытым слоем, состоявшим из 21 нейрона. В качестве функции потерь использовалась:

$$-(y \cdot \log(p) + (1 - y) \cdot \log(1 - p)),$$

где $y \in \{0, 1\}$ — истинное значение плана, $p \in \{0, 1\}$ — оценка вероятности принадлежности плана к заданному классу, вычисленная нейронной сетью.



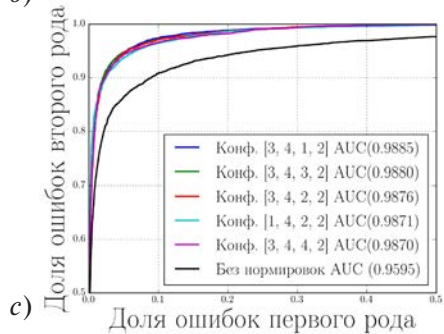
а)



б)

Рис. 4. Пять лучших конфигураций нормировок:

а — компьютерная графика/остальные; б — конвертация/остальные; в — стереоскопическая съёмка/остальные



в)

На рис. 4 представлено сравнение точности классификации с использованием различных конфигураций нормировок и подхода без использования нормировки.

Для дальнейших экспериментов были выбраны конфигурации: [1, 4, 2, 2], [1, 4, 3, 2], [1, 4, 1, 2], [1, 4, 2, 3], [1, 4, 3, 1], [3, 4, 1, 2], [3, 4, 3, 2].

2.2. СРАВНЕНИЕ КЛАССИФИКАТОРОВ ДЛЯ МНОГОКЛАССОВОЙ КЛАССИФИКАЦИИ

При переходе к классификации на три класса были испробованы:

- одна нейронная сеть, классифицирующая планы на три класса;
- подход «один против всех»;
- подход «один против одного».

Подход «один против всех» заключается в обучении трёх нейронных сетей, разделяющих конвертацию/съёмку, конвертацию/графику и съёмку/графику. Ответом будет являться результат классификатора, имеющего максимальную вероятность. Подход «один против одного» состоит в том, что обучаются три нейронные сети, разделяющие конвертацию/съёмку, конвертацию/графику и съёмку/графику. Результатом анализа будет тот класс, за который проголосовало большее количество классификаторов.

Все три подхода сравнивались на одинаковой обучающей выборке, для корректной оценки качества классификации использовалась кросс-валидация. Для увеличения корректности результатов обучение и валидация производились на разных фильмах. Результаты проведенного тестирования приведены на рис. 5:

- в конфигурациях слоёв нейронных сетей указаны только скрытые слои;
- FC(N) — полносвязный скрытый слой, содержащий N нейронов;
- D($X\%$) — Dropout [6] слой, X — вероятность исключения нейронов;
- подход «один против всех» — использовалось три нейронных сети с конфигурациями:

1) графика/остальные: слои [FC(33), D(10%)], нормировки [3, 4, 3, 2];

2) конвертация/остальные: слои [FC(10), D(15%), FC(5), D(10%)], нормировки [1, 4, 3, 2];

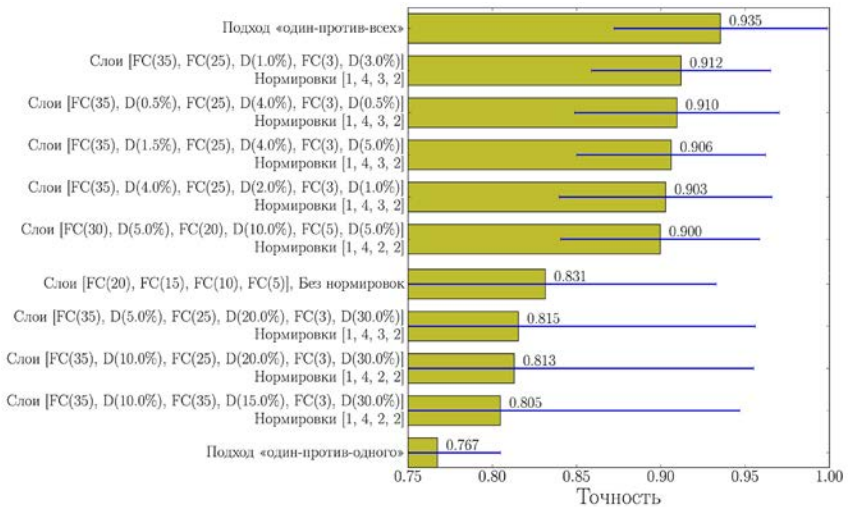


Рис. 5. Сравнение различных конфигураций

3) съёмка/остальные: слои [FC(33)], нормировки [3, 4, 3, 2];

— подход «один против одного» — использовались три нейронных сети с конфигурациями:

1) конвертация/съёмка: слои [FC(25)], нормировки [3, 4, 3, 2],

2) графика/конвертация: слои [FC(20)], нормировки [1, 4, 4, 2],

3) графика/съёмка: слои [FC(33)], нормировки [2, 4, 3, 1].

В таблице 1 приведена матрица ошибок для подхода «один против всех», показавшего лучшие результаты на обучающей выборке.

Таблица 1

Объективная оценка качества классификации лучшей конфигурации

Истинные значения	Предсказанные значения		
	Графика	Конвертация	Съёмка
Графика	95,81%	1,09%	3,08%
Конвертация	9,43%	89,52%	1,03%
Съёмка	3,03%	1,07%	95,89%

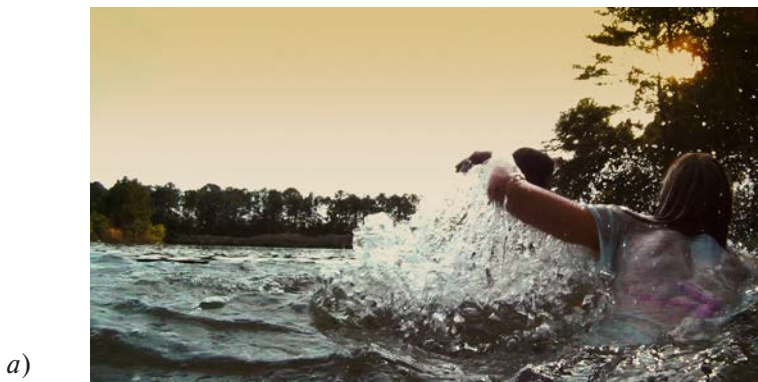


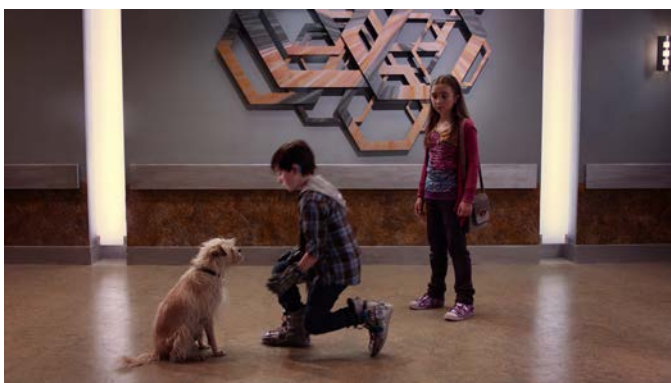
Рис. 6. Кадр найденного сконвертированного плана из фильма «Пиратанья 3D»:

a — левый кадр; *б* — правый кадр; *с* — анаглиф

а)



б)



в)



Рис. 7. Кадр найденного снятого плана из фильма «Дети шпионов 4D»: а — левый кадр; б — правый кадр; в — визуализация цветового искажения

2.3. АНАЛИЗ ПОЛНОМЕТРАЖНЫХ ФИЛЬМОВ

Предложенным алгоритмом было проанализировано 105 полнометражных стереоскопических фильмов, среди которых было 65 фильмов, созданных методом стереоскопической съёмки, и 40 фильмов, созданных с помощью конвертации.

В результате обнаружено 456 планов с конвертацией и 86 планов со съёмкой в снятых и сконвертированных фильмах соответственно.

Пример найденного сконвертированного плана изображён на рис. 6, снятого — на рис. 7. Все найденные планы опубликованы в 9-ом отчёте проекта VQMT3D [7].

ЗАКЛЮЧЕНИЕ

В данной работе рассмотрена модификация алгоритма, предложенного в [1], позволившая увеличить точность классификации на выборке из 16 фильмов с 83,14% до 93,57%.

Проведён анализ 105 полнометражных стереоскопических фильмов, среди которых найдено 456 планов с конвертацией в 65 снятых фильмах и 86 планов со съёмкой в сконвертированных фильмах.

ЛИТЕРАТУРА

1. Звездаков С.В., Боков А.А., Ватолин Д.С. Алгоритм определения способа производства планов стереофильмов // Запись и воспроизведение объёмных изображений в кинематографе и других областях: VII Международная научно-практическая конференция, Москва, 23–25 апреля 2015 г.: Материалы и доклады. 2015. С. 238–248.

2. Au cinéma, vous préférez la version 2D ou 3D?

http://www.lesnumeriques.com/legrandforum/avis/Ecrans-TV-Videoprojecteurs/divers/preferez-version-cinema-sujet_14090_1.htm

3. Fischler M., Bolles R. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography // Communications of the ACM. 1981. Vol. 24. P. 381–395.

4. Kakaletsis E., Nikolaidis N. A technique for fake 3d (2d-to-3d converted) video recognition // Signal Processing Conference (EUSIPCO), 23rd European/ 2015. P. 106–109.

5. MSU 3D-video Quality Analysis

<http://compression.ru/video/vqmt3d/>

6. Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting // The Journal of Machine Learning Research. 2014. Vol. 15. P. 1929–1958.

7. VQMT3D Project: Report 9 on 3D-video Quality Analysis
<http://www.compression.ru/video/vqmt3d/report9/>