

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМЕНИ Н.Э. БАУМАНА  
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)

# Метод распознавания спортивных действий человека на видео с использованием локального дескриптора

Студент: Сысоева Виктория Романовна, ИУ7-82Б

Научный руководитель: Мальцева Диана Юрьевна

Москва, 2023

# Цель и задачи

**Цель:** разработать метод распознавания спортивных действий человека на видео с использованием локального дескриптора.

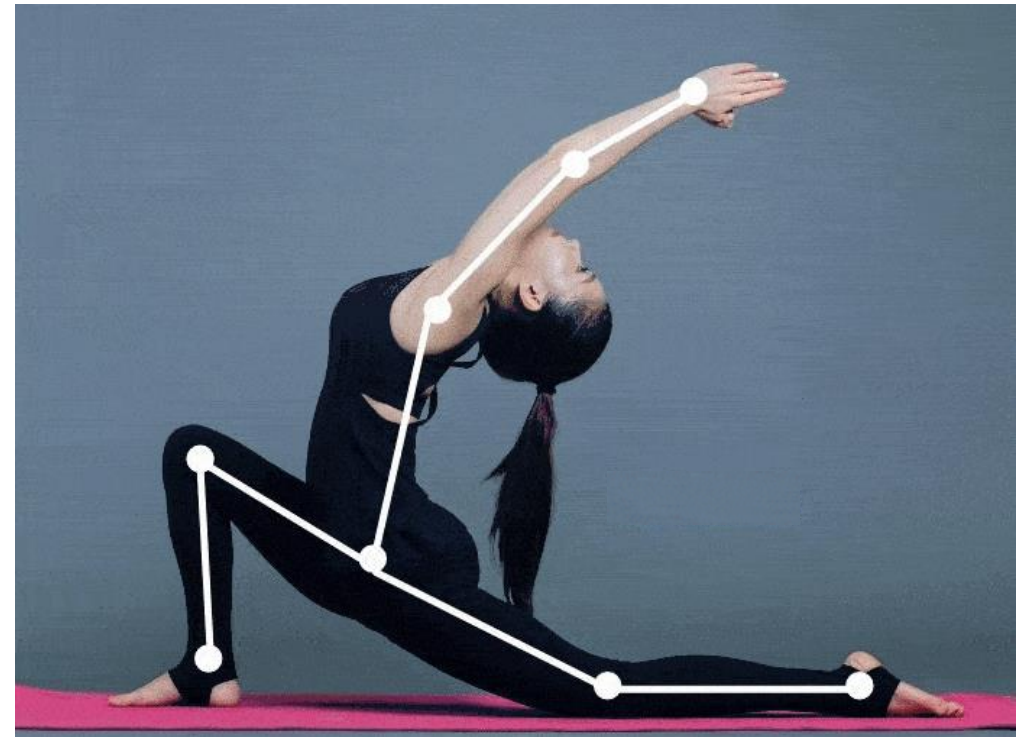
**Задачи:**

- Рассмотреть алгоритмы, с помощью которых можно реализовать метод распознавания спортивных действий человека.
- Сделать обзор существующих решений распознавания действий человека.
- Разработать метод распознавания спортивных действий человека на видео с использованием локального дескриптора.
- Спроектировать и реализовать ПО, демонстрирующее работу метода.
- Исследовать метрики разработанного метода в зависимости от разметки данных.
- Исследовать разработанный метод на применимость при работе с видео различной контрастности и качества.

# Области применения

Области применения разработанного метода распознавания спортивных действий человека:

- 1) Медицина (дистанционные тренировки при реабилитации)
- 2) Образование (ФВ МГТУ им. Н.Э. Баумана)



# Методы распознавания действий человека

Категория	Рассмотренный метод	Метод машинного обучения	Модальность данных	Набор данных	Точность
<b>Пространственно-временные</b>	Соомро К., Шах М.	Метод опорных векторов	RGB	UCF-Sports, Sub-JHMDB, JHMDB, THUMOS13	63,2% 45,9% 67,3% 54,4%
	Х. Идресс, М. Шах	Метод опорных векторов	RGB	UCF-Sports, sub-JHMDB THUMOS13	55% 42% 48%
Стохастические	Тахир С., Догар А.	Метод случайного леса	RGB	IM-WSHA, PAMAP-2, UCI HAR, MobiAct	90,18% 91,25% 91,83% 90,46%
	Ю. А. Егорова, И. Г. Захарова, А. Р. Гасанов, А. А. Филицих	Скрытая Марковская модель, сверточная нейронная сеть	RGB, скелетная модель	MPII Human Pose	75%
Основанные на правилах	Шарма К. М., Кушваха А.	GMM модель, нейронная сеть	RGB	KTH, набор данных Вейцмана	97,2% 91%
	Пак, Дж. Джин, Й. Чо, С.	Скрытая Марковская модель, FNN	RGB, скелетная модель	LALR-1	80,5%
Основанные на форме	Ван, Й., Мори, Джи	Марковские случайные поля	RGB	KTH, набор данных Вейцмана	73% 88%
	Икизлер Н., Дуйгулу П.	Метод опорных Векторов, сверточная нейронная сеть	RGB	Набор из 8 действий	95%

# Алгоритмы детектирования объектов с помощью дескриптора

Название алгоритма	Вычислительная сложность*	Устойчивость к шуму	Проприетарность	Инвариантность к масштабу	Устойчивость к освещению
SIFT (Scale Invariant Feature Transform)	-	-	-	+	-
SURF (Speeded Up Robust Features)	-	+	-	+	-
RIFF (Rotation Invariant Fast Features)	-	+	+	+	+
<b>HOG (Histogram of Oriented Gradients)</b>	+	+	-	+	-

\* Низкая вычислительная сложность обозначена символом «+», высокая «-».

В качестве метода детектирования был выбран HOG дескриптор.

# Методы классификации

Метод классификации	Устойчивость к выбросам	Ограничение на формат данных	Высокая алгоритмическая распараллеливаемость	Алгоритмическая сложность классификации
Метод Байеса	Да	Да	Нет	$O(cd)$
Дерево решений	Нет	Нет	Да	$O(\log n)$
<b>Случайный лес</b>	Да	Нет	Да	$O(k * \log n)$
Метод опорных векторов	Да	Да	Нет	$O(nd)$
К-ближайших соседей	Нет	Да	Нет	$O(nd)$

В качестве метода классификации был выбран метод случайного леса.

# Классификация видов спортивных действий

Распознается 10 видов действий:

1) мостик  
(108 видео)



2) поднятие рук  
(330 видео)



3) джампинг Джек  
(123 видео)



4) выпады  
(211 видео)



5) подтягивание  
(178 видео)



6) отжимание  
(139 видео)



7) пресс  
(146 видео)



8) приседание  
(122 видео)



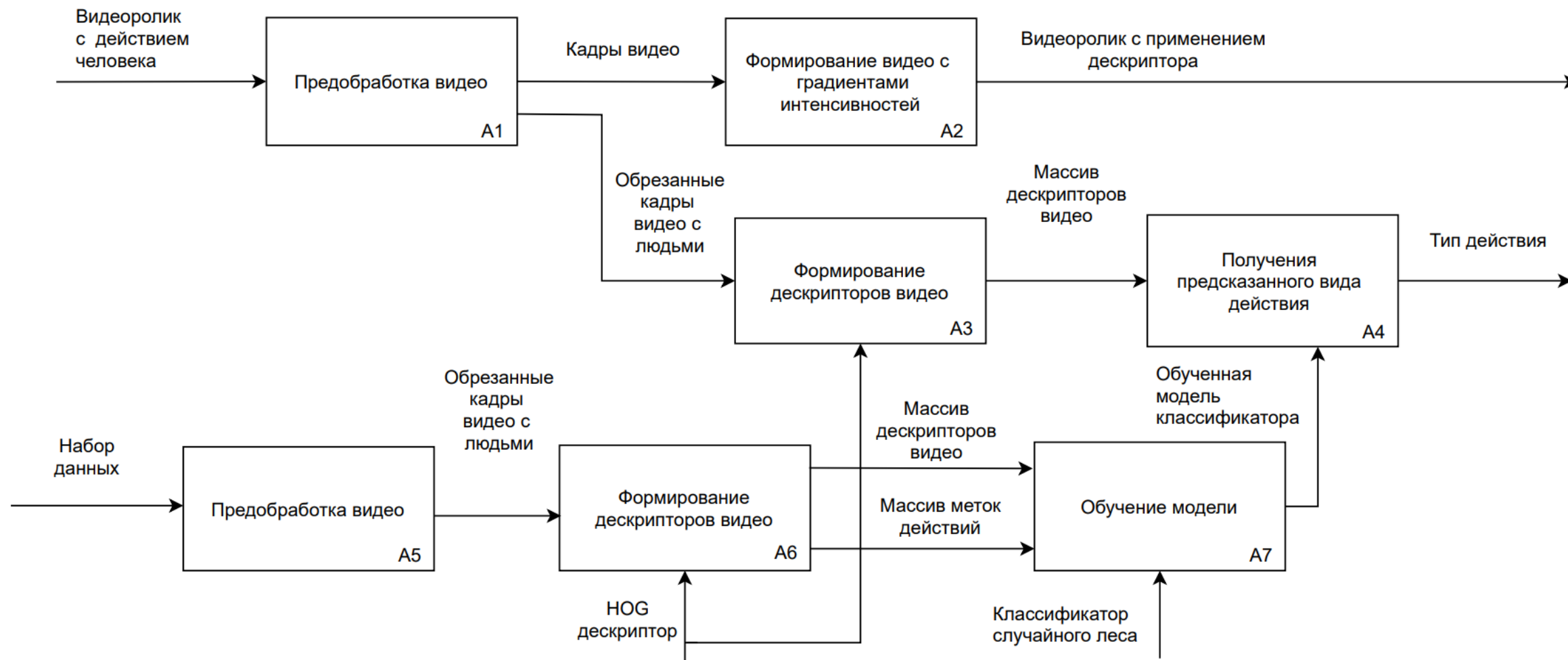
9) растяжка ног  
(154 видео)



10) махи ногами  
(129 видео)

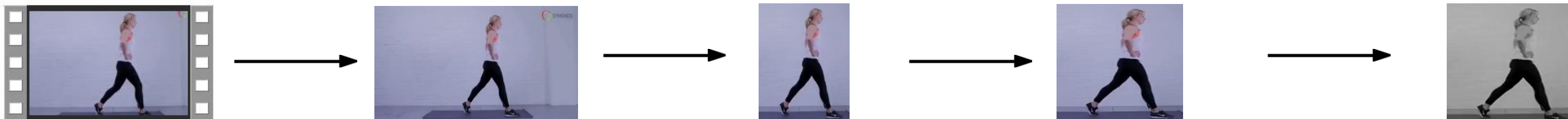
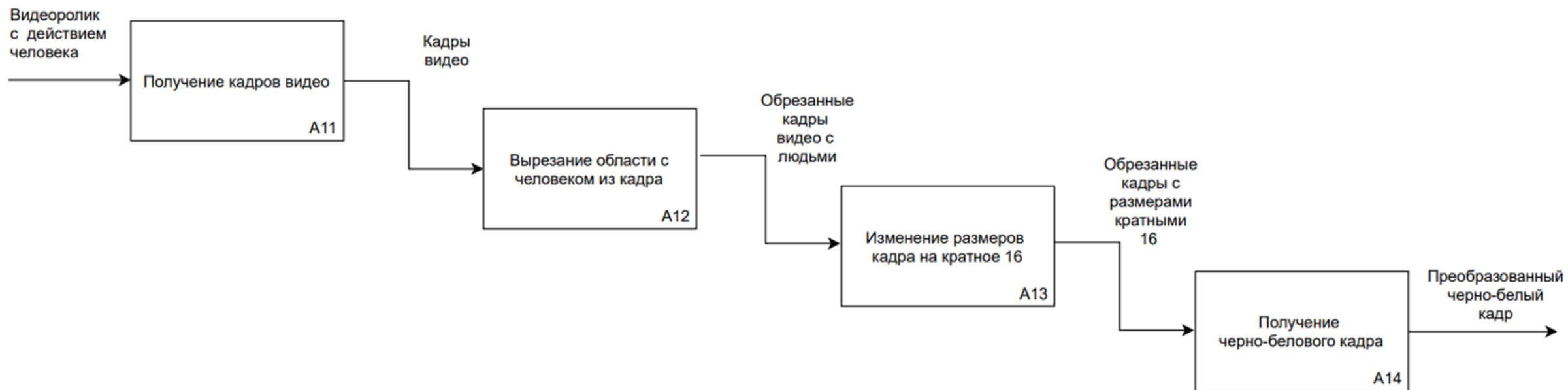


# Метод распознавания спортивных действий человека на видео





# Предобработка видео



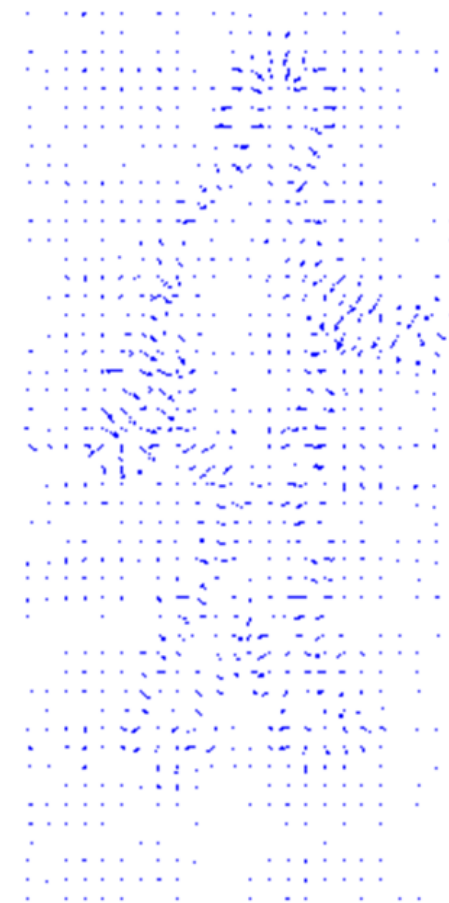
# Алгоритм HOG ( 1/3 )



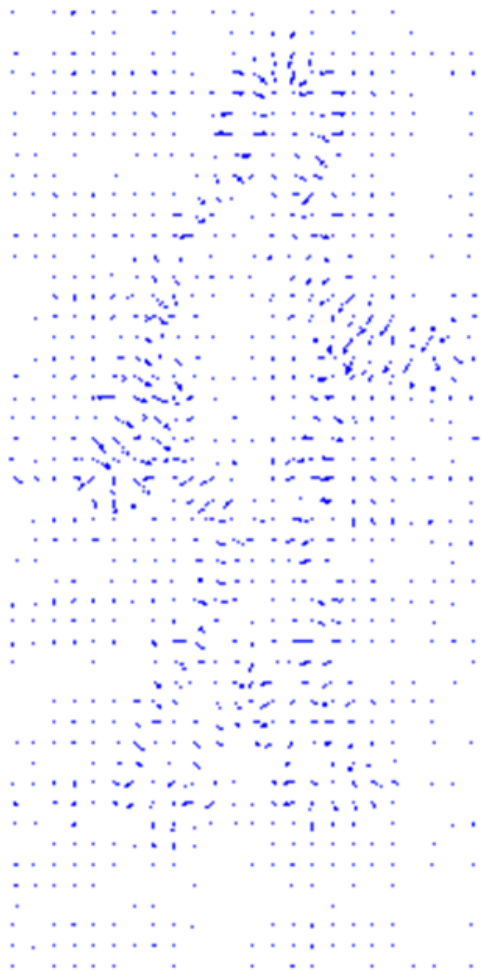
Чтобы вычислить дескриптор HOG, нужно вычислить горизонтальный и вертикальный градиенты  $G_x$  и  $G_y$ .

Затем вычисляется величина и направление градиента, используя следующие уравнения:

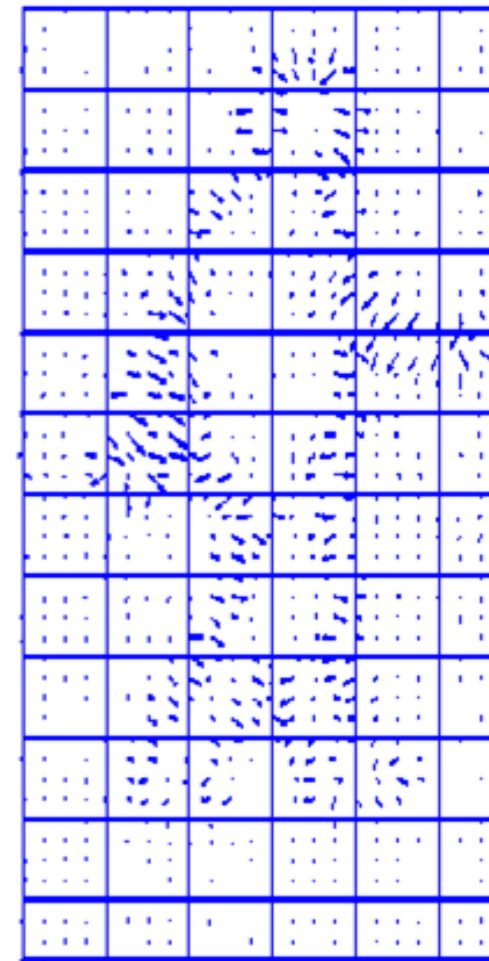
$$M(magnitude) = \sqrt{G_x^2 + G_y^2}$$
$$D(direction) = \arctan \frac{G_x}{G_y}$$



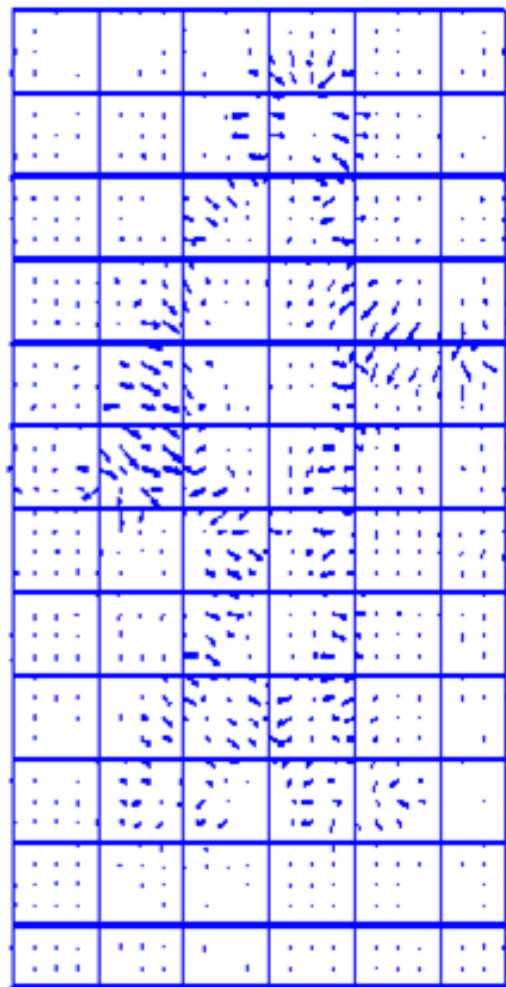
# Алгоритм HOG ( 2/3 )



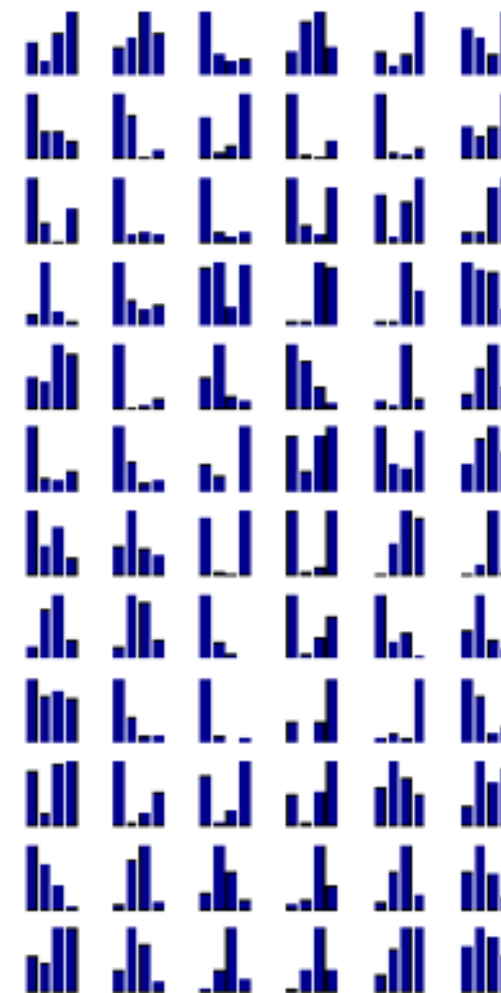
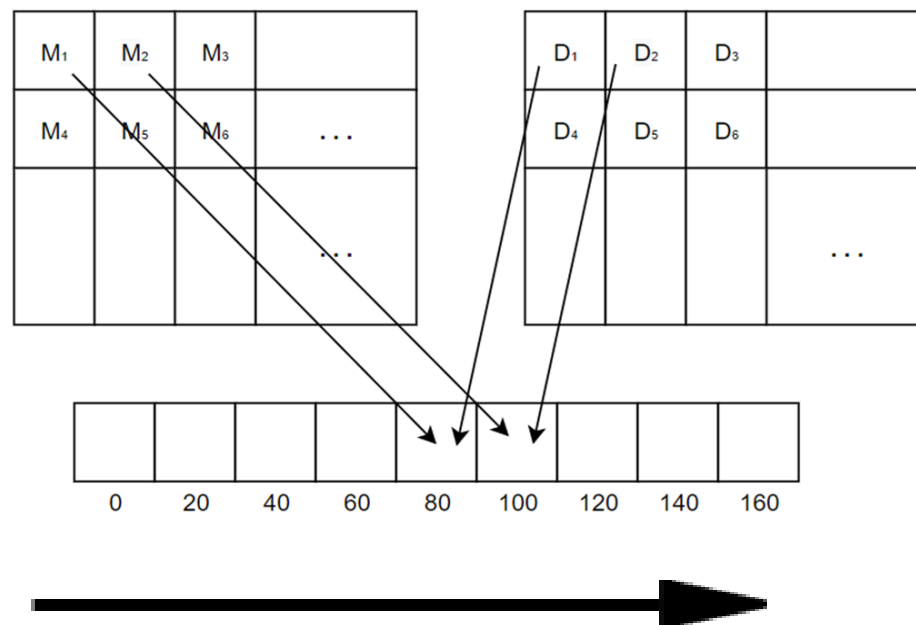
На этом этапе изображение делится на ячейки  $8 \times 8$  или  $16 \times 16$ . Вычисление гистограммы по фрагменту делает это представление более устойчивым к шуму.



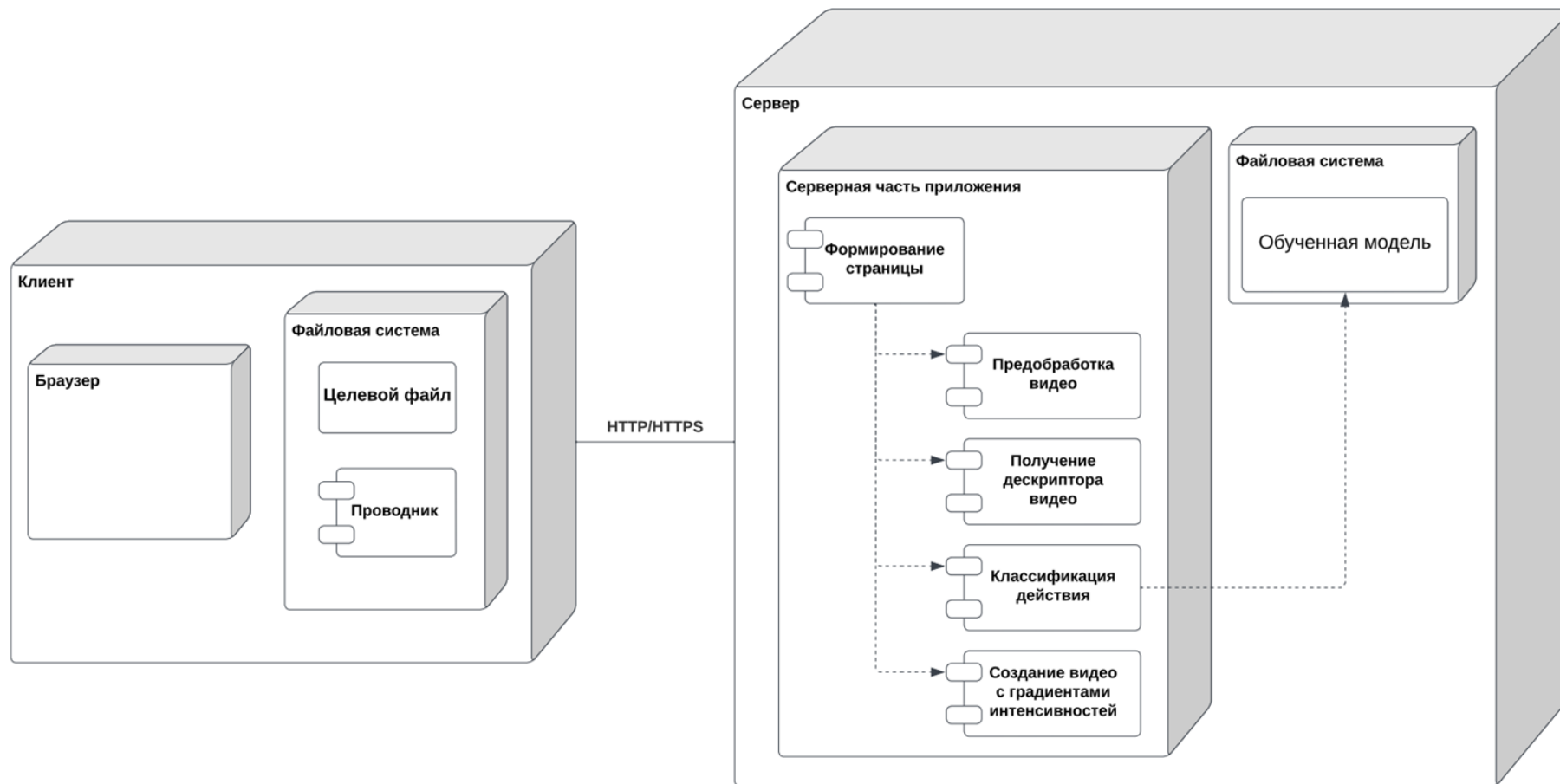
# Алгоритм HOG ( 3/3 )



Гистограмма градиентов вычисляется для каждой ячейки, на которые разделено изображение. Гистограмма содержит 9 интервалов, соответствующих углам 0, 20, 40 ... 160. Ниже показан процесс её формирования.



# Схема разработанного ПО

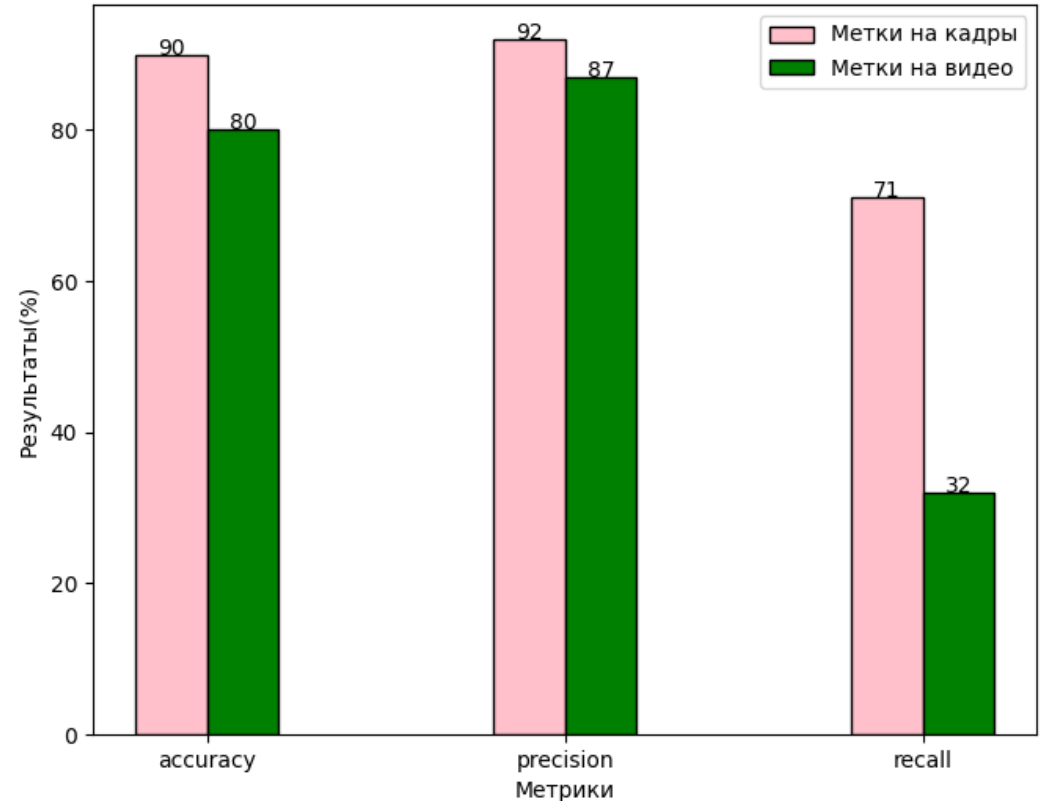


- Python
- Flask
- OpenCV
- Scikit-learn

# Качество метрик модели в зависимости от разметки данных

Матрица ошибок

Мостик	0.54	0.04	0.13	0.05	0.04	0.02	0.07	0.05	0.05	0.02
Приседания	0.01	0.61	0.17	0.02	0.04	0.04	0.04	0.04	0.01	0.02
Поднятие рук	0.01	0.04	0.56	0.08	0.08	0.13	0.0	0.06	0.02	0.02
Джампинг джэк	0.02	0.01	0.02	0.86	0.05	0.01	0.01	0.02	0.0	0.0
Выпады	0.02	0.02	0.18	0.03	0.55	0.08	0.02	0.02	0.05	0.03
Подтягивание	0.02	0.02	0.12	0.0	0.03	0.68	0.04	0.04	0.03	0.02
Отжимания	0.01	0.01	0.05	0.06	0.02	0.04	0.65	0.1	0.05	0.01
Пресс	0.04	0.01	0.07	0.01	0.03	0.0	0.09	0.66	0.09	0.0
Растяжка ног	0.03	0.01	0.11	0.01	0.06	0.05	0.05	0.01	0.63	0.03
Махи ногами	0.02	0.0	0.05	0.01	0.05	0.03	0.04	0.02	0.04	0.73

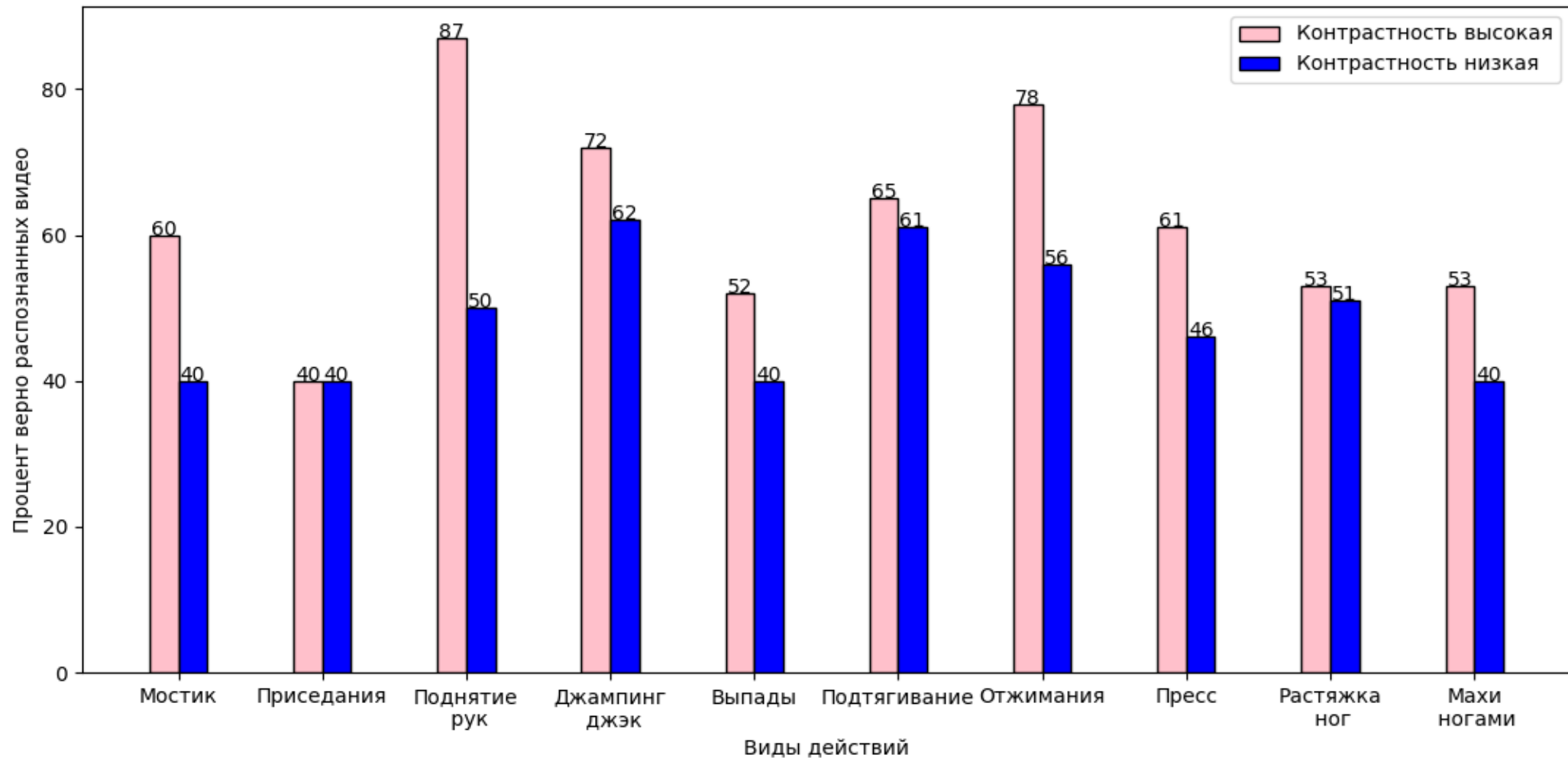


$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP}$$

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

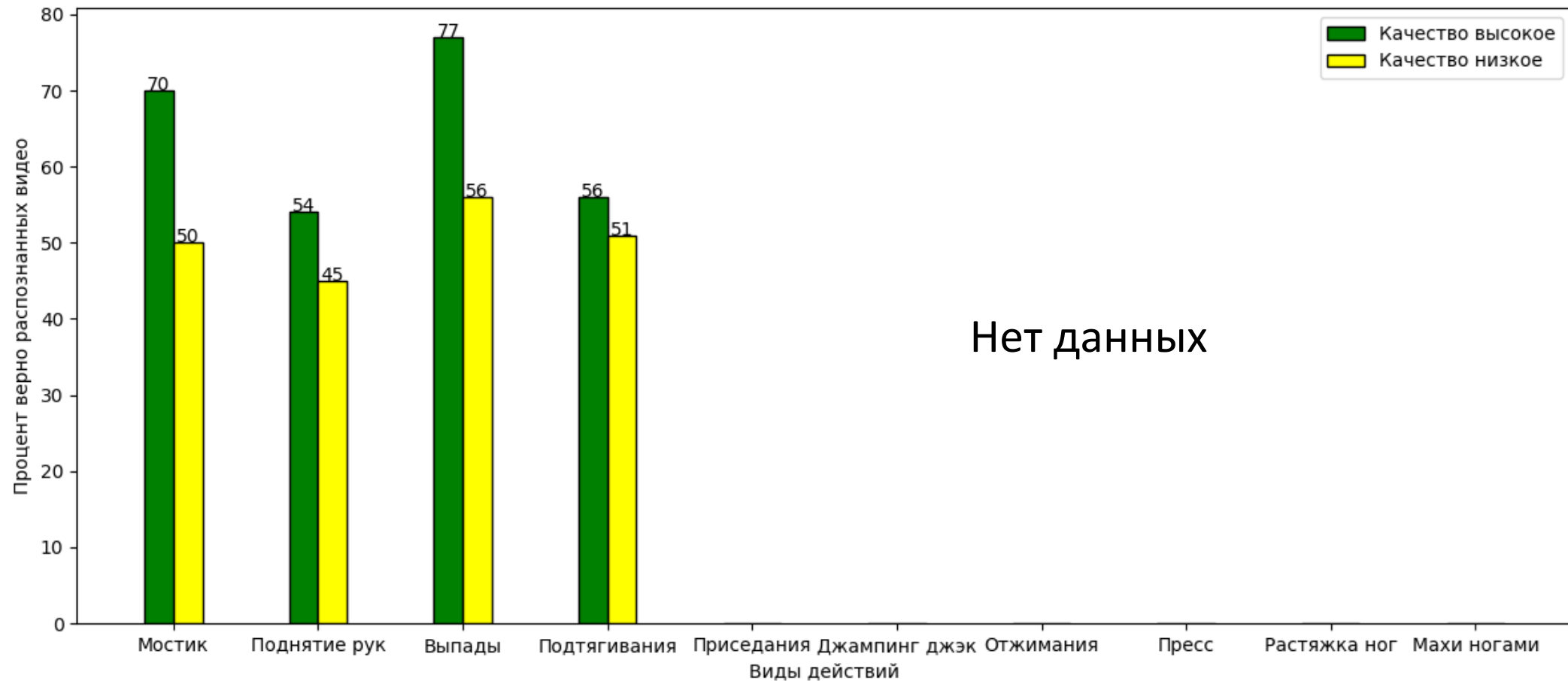
# Зависимость доли распознанных движений от контрастности видео



# Зависимость доли распознанных движений от качества видео

Высокое качество > 256x256 пикселей

Низкое качество <= 256x256 пикселей





# Выводы

В результате проведенных исследований можно сделать вывод о том, что точность метода выше при назначении метки на дескриптор каждого кадра, нежели при назначении на дескриптор видео. Также у рассмотренных методов, относящихся к пространственно-временной категории, точность составляла 42% – 67,3% в зависимости от набора данных, на котором проводилось тестирование, при этом разработанный метод имеет точность 92%.

# Заключение

Был разработан метод распознавания спортивных действий человека на видео с использованием локального дескриптора.

Были решены задачи:

- Рассмотрены алгоритмы, с помощью которых можно реализовать метод распознавания спортивных действий человека.
- Сделан обзор существующих решений распознавания действий человека.
- Разработан метод распознавания спортивных действий человека на видео с использованием локального дескриптора.
- Спроектирован и реализовать ПО, демонстрирующее работу метода.
- Проведено исследование метрик разработанного метода в зависимости от разметки данных.
- Проведено исследование разработанного метода на применимость при работе с видео различной контрастности и качества.

# Дальнейшее развитие и публикации

- Распознавание действий нескольких людей.
- Распознавание нескольких типов действий на одном видео.
- Автоматическое обрезание области с человеком из кадра.

Принята к печати публикация на тему «Метод распознавания спортивных действий человека на видео с использованием локального дескриптора» на студенческой конференции «Студенческая научная весна» (секция ИУ-7), 2023 г. (РИНЦ)