

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ «МЭИ»

На правах рукописи



Рыбинцев Андрей Владимирович

**ИССЛЕДОВАНИЕ, МОДИФИКАЦИЯ И РАЗРАБОТКА
МЕТОДОВ КОМПЬЮТЕРНОГО ЗРЕНИЯ
ДЛЯ ЗАДАЧ ОПРЕДЕЛЕНИЯ АТТРИБУТОВ ЛИЧНОСТИ
ПО ИЗОБРАЖЕНИЮ ЛИЦА**

Специальность 05.13.17 – Теоретические основы информатики

Диссертация на соискание ученой степени
кандидата технических наук

Научный руководитель:
д.т.н., профессор
Фролов Александр Борисович

ОГЛАВЛЕНИЕ

Список сокращений	5
ВВЕДЕНИЕ.....	6
1. АНАЛИЗ СУЩЕСТВУЮЩИХ ПОДХОДОВ К ОПРЕДЕЛЕНИЮ АТТРИБУТОВ ЛИЧНОСТИ ПО ИЗОБРАЖЕНИЮ ЛИЦА	18
1.1. Место задачи определения атрибутов личности по изображению лица в общей проблеме распознавания образов	18
1.2. Актуальность задачи определения атрибутов личности по изображению лица	21
1.3. Основные этапы определения атрибутов личности по изображению лица	25
1.4. Задача классификации изображений как задача обучения по прецедентам	28
1.5. Анализ методов формирования вектора признаков изображения лица для определения атрибутов личности	33
1.6. Оценка качества методов определения атрибутов личности	47
1.7. Анализ методов классификаций изображений лиц людей по различным атрибутам	52
1.7.1. Методы классификации изображений по атрибуту «пол»	52
1.7.2. Методы классификации изображений по атрибуту «раса»	68
1.7.3. Методы классификации изображений по атрибуту «возраст» ...	69
1.8. Итоги анализа методов определения атрибутов личности по изображению лица	77
1.9. Выводы по главе 1	79

2. РАЗРАБОТКА МЕТОДА ПОСЛЕДОВАТЕЛЬНОГО ОПРЕДЕЛЕНИЯ АТТРИБУТОВ ЛИЧНОСТИ ПО ИЗОБРАЖЕНИЮ ЛИЦА	81
2.1. Формализованное описание локальных бинарных шаблонов на базе принципа конечной топологии	81
2.1.1. Принцип конечной топологии	81
2.1.2. Топологические формы изображений	83
2.2. Снижение размерности пространства признаков изображения для определения атрибутов личности.....	85
2.3. Модификация метода опорных векторов для бинарной классификации изображений при определении атрибутов «пол» и «раса».....	93
2.4. Модификация двухэтапной схемы определения атрибута «возраст» методом восстановления регрессии на основе опорных векторов.....	100
2.5. Метод последовательного определения атрибутов личности по изображению лица	109
2.6. Выводы по главе 2.....	114
3. ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ РАЗРАБОТАННОГО МЕТОДА ПОСЛЕДОВАТЕЛЬНОГО ОПРЕДЕЛЕНИЯ АТТРИБУТОВ ЛИЧНОСТИ ПО ИЗОБРАЖЕНИЮ ЛИЦА	117
3.1. Методика исследования точности определения атрибутов.....	117
3.2. Исследуемые наборы данных	119
3.2.1. Существующие наборы данных с изображениями лиц	119
3.2.2. Формирование набора данных изображений лиц по открытым источникам в сети Интернет	121

3.3. Результаты исследования точности разработанных модификаций и метода последовательного определения атрибутов личности	123
3.3.1. Точность определения атрибута «пол».....	123
3.3.2. Точность последовательного определения атрибута «раса» после определения атрибута «пол»	125
3.3.3. Точность определения атрибута «возраст» после определения атрибутов «пол» и «раса».....	127
3.4. Исследование затрат времени на определение атрибутов личности по изображению лица разработанным методом	129
3.5. Программная реализация разработанного метода последовательного определения атрибутов личности по изображению лица	131
3.6. Выводы по главе 3.....	131
ЗАКЛЮЧЕНИЕ	133
Литература.....	136
Приложение	149

Список сокращений

<i>Обозначение</i>	<i>Пояснение</i>
AAM	<i>Active Appearance Model</i> – активная модель внешности
ASM	<i>Active Shape Model</i> – активная модель формы
CS(ϵ)	<i>Cumulative Score</i> – суммарный счет с порогом ϵ (определяет долю объектов тестовой выборки, для которых значение возраста было определено с погрешностью меньше ϵ)
LBP	<i>Local Binary Patterns</i> – локальные бинарные шаблоны
LDA	<i>Linear Discriminant Analysis</i> – линейный дискриминантный анализ
MAE	<i>Mean Absolute Error</i> – средняя абсолютная ошибка (определяет среднюю абсолютную ошибку определения возраста по тестовой выборке)
SVM	<i>Support Vector Machine</i> – метод опорных векторов
SVR	<i>Support Vector Regression</i> – регрессия на основе опорных векторов
HOG	<i>Histogram of Oriended Gradients</i> – гистограмма ориентированных градиентов
EOH	<i>Edge Orientation Histogram</i> – гистограмма ориентаций границ
Accuracy	Метрика, определяющая долю верно классифицированных объектов тестовой выборки

ВВЕДЕНИЕ

Интерес к задаче автоматического определения различных атрибутов человека, к которым относятся пол, раса, возраст, выражение эмоций и пр., методами компьютерного зрения на основе анализа изображения лица в последние годы неуклонно растёт. Одной из главных причин этого является стремительное увеличение числа изображений людей в Интернете, и необходимость их автоматической аннотации для задач поиска. Поскольку человек является одним из наиболее распространённых объектов, встречающихся на фотографиях, количество изображений с людьми чрезвычайно велико и продолжает неуклонно расти. Автоматическое определение указанных атрибутов может помочь более качественно фильтровать результаты поиска.

Другим примером применения автоматического определения перечисленных выше атрибутов людей является анализ аудитории при проведении различных мероприятий. В частности, это полезно для формирования эффективной рекламы, которая должна быть адаптирована к половой, расовой и возрастной принадлежности аудитории. Традиционные социологические методы не могут обеспечить необходимую точность определения атрибутов за разумное время (перед началом мероприятия), тогда как автоматизированные системы, основанные на методах компьютерного зрения, могут учесть практически каждого человека и помочь докладчикам адаптироваться к особенностям аудитории.

Автоматическое определение атрибутов личности по изображению лица начинает применяться и при создании адаптивных человеко-машинных интерфейсов, в которых поведение системы подстраивается под конкретного пользователя, ориентируясь, в числе прочего, на его пол, возраст, эмоциональную реакцию.

Для определения атрибутов личности по изображению лица применяются различные методы автоматической классификации, т.е. методы отнесения изображений к заранее определенным категориям, на основе машинного обучения. Целью применения методов машинного обучения является построение модели классификации на основе обучающей выборки (обучение с учителем), с последующим применением данной модели для предсказания класса нового изображения.

В настоящее время системы автоматической классификации людей по атрибутам активно развиваются. Так, в частности, уже существуют системы, определяющие пол человека и расу с точностью, достаточной для практического применения.

Задача классификации людей по возрасту является наиболее востребованной и заведомо более сложной. Даже человеку не всегда удаётся определить возраст другого человека с приемлемой точностью, что накладывает дополнительные трудности на разработку алгоритмов классификации. Необходимо отметить, что при определении возраста в ряде случаев достаточно определить значение возраста приближенно, чтобы отнести человека к какой-то возрастной группе (подростки, юноши/девушки, люди средних лет, пенсионеры), но существуют и задачи, требующие определения более точного значения данного атрибута.

Основной проблемой в задаче классификации атрибутов является огромное разнообразие вариантов их проявления у разных людей. Так, например, люди одинакового возраста могут выглядеть совершенно по-разному, даже если они являются представителями одной расы и одного пола, вследствие индивидуальных особенностей и разного проявления процесса старения. Аналогичная ситуация имеет место и с другими атрибутами личности.

Другой проблемой являются сами изображения людей. Наибольшее количество информации о человеке можно получить из фронтального изображения его лица, поэтому преимущественно именно такие изображения применяются в задачах определения атрибутов личности. Однако получение подобных изображений в реальной жизни является также непростой задачей.

Необходимо отметить, что для практического применения задача определения атрибутов человека должна решаться за весьма ограниченное время (сохранять диалоговый режим в задачах поиска в Интернете, успеть определить атрибуты за время прохождения человека мимо видеокамеры и пр.) Данное требование накладывает дополнительные ограничения на производительность методов классификации по атрибутам, т.е. при разработке подобных методов игнорирование их производительности может сделать предложенное решение абсолютно бессмысленным с практической точки зрения.

В рамках настоящей работы в качестве атрибутов личности рассматриваются пол, раса и возраст человека, так как классификация изображений людей именно по этим атрибутам имеет наибольшую практическую ценность. При этом определение конкретного значения атрибута «возраст» является наиболее важным. Кроме того, эти атрибуты представляют собой примеры из трех основных категорий атрибутов личности: бинарная характеристика (пол - мужской/женский), множественная нечисловая характеристика (раса - европеоидная, монголоидная, негроидная) и численное значение некоторого признака (значение возраста). Развитие методов классификации атрибутов указанных категорий может быть использовано для определения и других атрибутов человека: наличие очков, проявление эмоций, пропорции лица и пр.

Цель диссертационной работы. Целью диссертационного исследования является повышение качества определения атрибутов личности (пола, расы, возраста) по изображению лица на основе использования современных методов автоматической классификации и машинного обучения.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Провести анализ существующих способов формирования пространства признаков, описывающих изображение, и методов определения значений этих признаков, и выбрать из них такие, которые позволят решить задачу совместного определения искомым атрибутов личности - «пол», «раса», «возраст».

2. При различных способах формирования вектора признаков изображения лица, проанализировать известные методы классификации изображений лиц с целью выбора наиболее перспективных методов для их совершенствования с учетом затрат времени на процедуры обучения и классификации.

3. Разработать способы понижения размерности пространства признаков, описывающих изображение лица, за счет исключения незначимых для классификации признаков с целью повышения качества классификации и сокращения как времени классификации, так и времени обучения классификатора.

4. Произвести модификацию существующих автоматических методов бинарной классификации изображений лица по атрибуту «пол» с целью повышения качества (по используемым сегодня критериям) определения данного атрибута личности.

5. Произвести модификацию существующих автоматических методов множественной классификации изображений лица по атрибуту

«раса» с целью повышения качества (по используемым сегодня критериям) определения данного атрибута личности.

6. Произвести модификацию существующих методов восстановления регрессии для определения атрибута «возраст» с целью повышения качества (по используемым сегодня критериям) определения данного атрибута личности.

7. Разработать метод последовательной автоматической классификации изображений лица для определения выбранных атрибутов личности («пол», «раса» и «возраст»), поэтапно учитывающий результаты предыдущих шагов классификации.

8. Провести анализ разработанного метода последовательной автоматической классификации изображений лиц людей для определения выбранных атрибутов личности («пол», «раса», «возраст») на предмет повышения качества классификации по используемым в настоящее время критериям качества на базах реальных изображений людей.

Объект исследования. Объектом исследования является теория и практика машинного обучения.

Предмет исследования. Предметом исследования являются методы и средства машинного обучения для определения атрибутов личности по изображению лица.

Методы исследования. Полученные в диссертации результаты основываются на применении методов машинного обучения, распознавания образов, цифровой обработки изображений, теории вероятностей, математической статистики, методов нелинейной оптимизации, линейной алгебры и математического анализа.

Достоверность полученных результатов. Достоверность полученных научных результатов подтверждается их согласованностью с опубликованными научными данными, полученными российскими и зарубежными исследователями, и компьютерными экспериментами.

Научная новизна. Основные результаты работы являются новыми и заключаются в следующем:

1) На основе анализа существующих методов признакового описания изображений лиц людей для формирования вектора признаков обосновано использование метода локальных бинарных шаблонов (*LBP*), что позволяет сформировать пространство признаков большой размерности (несколько тысяч), обеспечивая высокую концентрацию информации об исходном изображении, и создает предпосылки для более точной классификации по возрасту.

2) Разработан метод снижения размерности пространства признаков изображений, включающий автоматическую процедуру определения значимости каждого признака на базе алгоритма *Adaboost*, который позволил, снизить размерность пространства признаков изображения почти в 10 раз по сравнению со стандартным методом *LBP*.

3) Предложена модификация метода бинарной классификации, отличающаяся использованием применявшейся ранее при решении других задач компьютерного зрения процедуры бутстраппинга (обучение на «трудных» примерах).

4) Процедура бутстраппинга усовершенствована путем исключения «самых трудных» примеров (*усеченный* бутстраппинг) для устранения эффекта возможного переобучения решающей функции с целью снижения влияния на ее формирование незначимого для классификации «шума» в обучающей выборке.

5) Предложена модификация метода множественной классификации, как последовательности бинарных классификаций, отличающаяся использованием предложенной модифицированной *усеченным* бутстраппингом бинарной классификации.

6) Предложена модификация двухэтапной схемы определения атрибута «возраст» методом восстановления регрессии, отличающаяся

использованием на втором этапе зависимости чувствительности функции потерь и диапазона возрастов для построения регрессии от результатов первого этапа. При этом в процессе формирования решающей функции для определения возраста используется разработанная процедура *усеченного бутстраппинга*.

7) Разработан метод последовательного определения атрибутов личности («пол» → «раса» → приближенное значение атрибута «возраст» → уточненное значение атрибута «возраст»), с использованием перечисленных выше модификаций, и отличающийся учетом на каждом следующем шаге результатов предыдущего.

Практическая ценность.

1. Реализация разработанного метода последовательного определения атрибутов личности по изображению лица при использовании открытых библиотек программ методов *SVM*, *SVR* и *Adaboost* позволила на существующих открытых базах изображений лиц людей и сформированной в рамках настоящей работы собственной базе изображений повысить качество распознавания (по критерию *Accuracy*) пола – на 13%, расы – на 15% и возраста по критерию *MAE* почти на 2 года, а *CS(5)* – на 20%. На базе реальных изображений лиц людей, полученных из сети Интернет, достигнуто качество классификации по критерию *Accuracy* пола – 92 %, расы, после определения пола, – 89% и значения возраста после определения пола и расы по критерию *MAE* – 6,09 лет.

2. Реализация метода снижения размерности пространства признаков, сформированного методом *LBP*, позволила сократить время обучения классификаторов и совместной классификации пола, расы и возраста в 8 раз при повышении качества классификации по используемым критериям определения атрибутов личности.

Результаты диссертационной работы используются в компании ООО «ДаблЭй», созданной по инициативе и при поддержке государственного фонда Содействия развитию малых форм предприятий в научно-технической сфере, (программа «Старт», грант С1-19968: «Разработка программного обеспечения для автоматической возрастной классификации информационной продукции»). Данный фонд (Фонд содействия инновациям) является государственной некоммерческой организацией в форме федерального государственного бюджетного учреждения, образованного в соответствии с постановлением Правительства Российской Федерации от 3 февраля 1994 г. № 65.

Результаты диссертации также отражены в отчете о НИР «Оценка сложности алгебраических преобразований и их реализация в криптографических протоколах и в системах распознавания образов». Гос. рег. № 01201456557 по теме № 3029140, 2016 г., выполненной по проекту РФФИ № 14-01-00671а.

Апробация. Положения диссертационной работы докладывались на международной конференции «Графикон-2012», на 22-ой международной научно-технической конференции студентов и аспирантов «Радиоэлектроника, электротехника, энергетика» в 2016 г., обсуждались на рабочих совещаниях компании «ДаблЭй», в Фонде содействия инновациям и на научном семинаре по дискретным математическим моделям на кафедре математического моделирования «НИУ «МЭИ».

Публикации. По теме диссертации опубликовано 8 работ, в том числе 5 в изданиях перечня ВАК (2 работы опубликованы в изданиях, индексируемых в Scopus и Web-of-Science).

Объем и структура работы. Диссертация состоит из введения, трех глав, заключения и приложения. Список используемой литературы содержит 126 наименований. Текст диссертации содержит 148 страниц машинописного текста, включая 19 рисунков, 11 таблиц и 1 приложение.

В первой главе обосновывается актуальность определения атрибутов личности по изображению лица, приводится общее описание технологии и осуществляется обзор методов поэтапного решения этой задачи с целью выявления направления исследований, наиболее перспективных для дальнейшего развития. Показано, что наиболее перспективным методом формирования вектора признаков является метод локальных бинарных шаблонов (*LBP*) как для задачи определения атрибутов «пол» и «раса», так и для атрибута «возраст», и что в качестве метода обучения (метода формирования решающей функции) целесообразно использовать метод опорных векторов (*SVM*) для определения атрибутов «пол» и «раса», и регрессию на основе опорных векторов (*SVR*) для определения атрибута «возраст». Оценку качества методов классификации при этом целесообразно осуществлять с помощью хорошо зарекомендовавших себя критериев: *Accuracy* при определении атрибутов «пол» и «раса», средняя абсолютная ошибка по тестовой выборке (*Mean Absolute Error (MAE)*) и суммарный счет с порогом ϵ (*Cumulative Score, CS(\epsilon)*) - при определении атрибута «возраст».

Во второй главе приводятся оригинальные результаты:

- 1) Метод снижения размерности пространства признаков изображений, формируемых методом локальных бинарных шаблонов (*LBP*), включающий автоматическую процедуру определения значимости каждого признака на базе алгоритма *Adaboost*, который позволил, за счет использования априорной информации о человеческом лице, снизить размерность пространства признаков изображения почти в 10 раз по сравнению со стандартным методом *LBP*;
- 2) Модифицированный метод бинарной классификации на основе опорных векторов (*SVM*), отличающийся использованием применявшейся ранее при решении других задач компьютерного

- зрения процедуры бутстраппинга (обучение на «трудных» примерах);
- 3) Усовершенствованная процедура бутстраппинга (обучение на «трудных» примерах), путем исключения «самых трудных» примеров (*усеченный* бутстраппинг) для устранения эффекта возможного переобучения решающей функции с целью снижения влияния на ее формирование незначимого для классификации «шума» в векторе признаков;
 - 4) Модифицированный метод множественной классификации как последовательности бинарных решений, отличающийся использованием предложенной модифицированной *усеченным* бутстраппингом бинарной классификации.
 - 5) Модификация двухэтапной схемы определения атрибута «возраст» методом восстановления регрессии на основе опорных векторов (*SVR*); в рамках данной схемы на первом этапе предварительно определяется значение атрибута «возраст» по регрессии, построенной по всей обучающей выборке (этап глобальной регрессии), а на втором этапе - строится локальная регрессия по выделенному вокруг приближенного значения возраста диапазону с «плавающими», а не фиксированными границами, значение которых зависит от возраста; при этом в ϵ -чувствительной функции потерь чувствительность ϵ также зависит от предварительно вычисленного, значения возраста, а не является постоянной, как в существующих подходах; в процессе формирования решающих функций для определения приближенного и уточненного значений возраста используется разработанная процедура *усеченного* бутстраппинга;
 - 6) Метод последовательного определения атрибутов личности («пол» → «раса» → приближенное значение атрибута «возраст» →

уточненное значение атрибута «возраст»), в котором на каждом следующем этапе учитываются результаты предыдущего и используются перечисленные выше модификации существующих методов для определения атрибутов личности, т.е. сформирована единая процедура совместного определения атрибутов личности, которая может быть остановлена на любом шаге в зависимости от специфики решаемой задачи.

Третья глава посвящена исследованию точности и трудоемкости разработанного метода с учетом предложенной процедуры формирования вектора признаков, и содержит результаты, полученные при решении поставленной в работе задачи определения атрибутов личности («пол», «раса» и «возраст») на существующих открытых базах изображений лиц людей. Поскольку существующие открытые базы изображений позволяют лишь частично решить задачу исследования точности разработанного подхода, в рамках настоящей работы была поставлена и решена задача формирования собственной базы изображений лиц людей на основе открытых источников в сети Интернет, а для получения достоверных оценок точности использована процедура k -кратного скользящего контроля. Показано, что предложенные модификации и разработанный метод последовательного определения атрибутов личности, позволяют по сравнению с лучшими существующими методами на 13% повысить качество классификации по атрибуту «пол», на 15% - качество классификации по атрибуту «раса», а при определении атрибута «возраст» - на 1,98 года по критерию MAE и на 20% по критерию $CS(5)$ при анализе реальных изображениях лиц, полученных из сети Интернет.

В заключении приведены основные результаты, полученные в настоящей работе, и предлагаются возможные пути дальнейшего развития предложенных подходов.

Приложение содержит Акт компании ООО «ДаблЭй» о внедрении результатов диссертационной работы.

1. АНАЛИЗ СУЩЕСТВУЮЩИХ ПОДХОДОВ К ОПРЕДЕЛЕНИЮ АТТРИБУТОВ ЛИЧНОСТИ ПО ИЗОБРАЖЕНИЮ ЛИЦА

1.1. Место задачи определения атрибутов личности по изображению лица в общей проблеме распознавания образов

Очередное исследование компании IDC показало, объем информации в мире растет примерно на 60% в год [29], и к 2020 году оценивается в 40 зеттабайт ($\sim 40 \cdot 10^{21}$). Важно отметить, что основная часть этих данных представлена в неструктурированном виде. В связи с этим задачу поиска и извлечения информации из столь стремительно растущих объемов можно решать только используя интеллектуальные системы [7], причем степень интеллектуальности подобных систем должна расти не медленнее, чем объем данных. Указанное обстоятельство постоянно стимулирует развитие существующих и разработку новых подходов к построению интеллектуальных систем, одним из базовых направлений при построении которых является распознавание образов. Интенсивное развитие именно этого направления связано с наличием мощного фундамента - математической теории распознавания и классификации [8, 9, 11, 19, 20, 21, 23, 24, 26, 27] как универсальной методологии математической обработки и оценивания информации с целью выявления скрытых закономерностей. Данная теория сформировалась благодаря работам целой плеяды выдающихся советских, российских и зарубежных ученых: Журавлева Ю.И., Матросова В.Л., Рудакова К.В., Вагина В.Н., Вапника В.Н., Ивахненко А.Г., Воронцова К.Н., Кудрявцева В.Б., Козлова В.Н., Неймана Дж., Пирсона Э., Вальда А., Розенблатта Ф., Фишера Р., Виолы П., Джонса М. и многих других.

С 80-х годов прошлого века одной из центральных парадигм интеллектуализации информационных систем стало машинное обучение

(*Machine Learning*), в процессе которого на компьютере пытаются воспроизвести процесс индуктивного обучения, присущий человеку, т.е. осуществляется группировка сходных объектов в классы и формирование правил, предсказывающих поведение объектов каждого класса [7].

В работах, посвященных созданию интеллектуальных систем на основе индуктивных методов, используются термины *классификация* и *распознавание*. Эти семантически близкие понятия на самом деле различны и приводят к двум разным подходам к обучению [7]: «обучение без учителя» и «обучение с учителем».

Обучение без учителя.

Данный подход предполагает [7, 11, 13], что интеллектуальной системе предъявляется совокупность объектов, не разделенных на классы (количество классов может быть неизвестным). Система должна сама определить классы, основываясь на общих свойствах предъявляемых объектов. Данная задача относится к задаче *распознавания*, под которой понимают формирование некоторого правила, по которому объекты, принадлежащие определенному классу, становится возможным отделить от остальных. Полученные классы могут быть взаимоисключающими (непересекающимися) или допускать более широкое представление в виде иерархии классов. В целом, данная задача называется задачей *кластеризации* [7, 11, 13].

Обучение с учителем.

Процедура обучения с учителем, предложенная Бонгардом [3], широко используется в системах машинного обучения и индуктивного формирования понятий. В обучении с учителем учитель определяет классы и предлагает примеры объектов каждого класса (обучающую выборку). Система должна найти общие свойства объектов каждого класса, сформировав тем самым описание классов. Описание классов вместе с самими классами формирует классифицирующее правило «Если

<описание> то <класс> », которое может быть использовано для предсказания класса объектов, неизвестных системе ранее [7]. Данная задача – это задача *классификации*: отнесение вновь предъявленных объектов, отличных от объектов обучающей выборки, к тому или иному классу на основе построенного классифицирующего правила. Полезность полученного правила проверяется на множестве контрольных примеров (тестовая выборка): правило, способное верно расклассифицировать только объекты исходной выборки, нельзя признать удачным. Таким образом, классификация изучает способы построения правила, способного отнести элемент данных к одному или нескольким известным классам (классифицировать элемент).

Среди неструктурированных данных сегодня, в связи с повсеместным внедрением мультимедийных систем, все большую долю занимают различные изображения. Поэтому, задачи распознавания и классификации изображений, являясь частью общей теории распознавания и классификации образов, становятся все более востребованными. При работе с исходной информацией, задаваемой в виде изображений, возникает целый ряд специфических задач, которые подробно рассмотрены в [22]:

1) задачи сопоставления:

сопоставление заданного изображения с набором изображений для установления их принадлежности к одному классу (определяется, представляют ли изображения один и тот же объект или сцену);

2) задача поиска:

поиск на предъявленном для распознавания изображении регулярности/нерегулярности/фрагмента заданного вида;

3) задача восстановления:

восстановление пропущенных кадров в последовательности изображений, восстановление изображений в целом по их фрагментам,

восстановление фрагментов изображения (и объектов) на основе производных элементов, признаков и порождающих процедур с учетом контекста изображения в целом;

4) задача классификации:

отнесение изображения к тому или иному классу.

Методы решения перечисленных задач, реализуемые с помощью интеллектуальных систем, принято относить к методам компьютерного зрения (*Computer Vision*) [10].

Развитие интернета и социальных сетей [97] привело к тому, что изображение человека вообще и его лица, в частности, стало одним самых распространенных неструктурированных графических объектов. Данное обстоятельство способствует все возрастающему интересу к разработке подходов к обработке именно изображений лиц людей с целью решения широкого круга задач. При этом наиболее практически востребованной является задача определения таких параметров (атрибутов) человека, как «пол», «раса» и «возраст».

Таким образом, настоящее исследование лежит в русле модификации существующих и разработки новых подходов к классификации изображений лиц людей с целью определения атрибутов личности («пол», «раса», «возраст») с помощью методов компьютерного зрения на базе машинного обучения в рамках общей теории распознавания образов.

1.2. Актуальность задачи определения атрибутов личности по изображению лица

Задача автоматического определения атрибутов личности по изображению лица на базе методов компьютерного зрения приобретает сегодня все большее значение, поскольку человек является наиболее часто

встречающимся объектом на фотографиях и видео. Основными атрибутами личности, определение которых имеет первостепенное практическое значение, являются пол человека, его раса и возраст [25, 28]. Среди практических задач, в которых именно эти атрибуты являются наиболее значимыми, следует выделить следующие:

- автоматическая нотация результатов поиска в Интернете;
- анализ аудитории при проведении маркетинговых мероприятий;
- формирование адаптивной наружной рекламы;
- создание адаптивного человеко-машинного интерфейса;
- контроль доступа личности к информации или определенным действиям.

Рассмотрим эти задачи более подробно.

На сегодня Интернет стал неотъемлемой частью нашей жизни, обойтись без которой не представляется возможным, а наличие целого ряда поисковых систем (Google, Yandex, Rambler, Mail.ru и пр.) существенно упрощает процесс извлечения требуемой информации. Поскольку изображения людей часто являются предметом поиска или могут помочь в поиске необходимой информации, качество выполнения этой операции (количество найденных в Интернете объектов, действительно удовлетворяющих запросу) чрезвычайно важно. Если, например, необходимо найти изображение Иванова Сергея, и известен год его рождения, то огромное количество изображений людей старшего возраста может быть отвергнуто поисковой системой, а если нужно найти изображение Сергиенко А.И. и известно, что это женщина, все изображения мужчин так же не войдут в результат поиска.

При проведении маркетинговых или иных публичных мероприятий для лектора чрезвычайно полезно перед началом выступления оценить состав аудитории, так как ее возрастная, половая и расовая принадлежность может существенно изменить набор используемых

лектором примеров, аналогий, цитат и шуток, что приведет к более эффективному воздействию на аудиторию. Сделать это человеку достаточно трудно, в то время, как автоматическая система, основанная на методах компьютерного зрения и установленная на входе в аудиторию, в состоянии учесть практически каждого человека. Понятно, что опытный лектор должен иметь «заготовки», ориентированные на различную аудиторию.

Наружная динамическая реклама (огромные экраны, воспроизводящие рекламные ролики) постепенно вытесняют стационарные плакаты, так как эффективность динамической рекламы намного выше. Объединение динамической рекламы с системой компьютерного зрения позволяет последней анализировать поток людей, проходящий рядом с таким экраном, с целью определения их атрибутов для выдачи наиболее значимой для данной категории лиц информации.

Поскольку компьютерные системы индивидуального обучения приобретают все большее распространение, чрезвычайно важно, чтобы процесс подачи учебного материала был максимально адаптирован к особенностям конкретной личности. При этом возраст, пол и даже раса человека могут иметь существенное значение для формирования темпа и порядка предоставления учебного материала [2].

Приведенные выше примеры не требуют точной классификации людей по возрасту. Для принятия правильного решения достаточно определить возраст приблизительно с точностью до 5-10 лет. Наряду с такими задачами существуют и другие, в которых требуется определить возраст с высокой точностью. Сегодня общепринята маркировка видеоматериалов по возрасту: 0+, 6+, 12+, 16+, 18+. Так, например, при автоматической продаже билетов на кинофильм категории 18+ целесообразно отказать 16-тилетнему подростку уже на этапе продажи билета, а не в момент его входа в кинозал. Подобный контроль необходим

и при доступе к видеоинформации, размещенной в Интернете (сегодня практически все устройства доступа имеют встроенную видеокамеру, так что технически такая задача вполне выполнима). При этом расовая принадлежность и пол личности могут существенно повлиять на результат. Так, возраст сексуального согласия [12] в различных странах зависит от пола и колеблется от 9 до 20 лет, а многие действия в странах Европы допустимы лишь по достижении 21 года [80]. Другим примером необходимости определения точного значения возраста является задача оценки продолжительности периода трудоспособности личности (до достижения пенсионного возраста). Важно, что результат такого анализа зависит от пола.

Понятно, что наряду с указанными атрибутами личности («пол», «раса», «возраст») практический интерес могут представлять и другие (например, проявление эмоций или наличие очков). Настоящее исследование ориентировано на анализ методов определения именно указанных трёх атрибутов как наиболее практически востребованных сегодня, но ее результаты могут быть использованы и для определения других атрибутов.

Отметим, что выделенные атрибуты личности относятся к трем важным категориям:

- бинарная (пол – «мужской», «женский»);
- множественная нечисловая (раса – «европеоидная», «монголоидная», «негроидная» [6]);
- числовая (возраст - числовое значение).

Поэтому для их определения в настоящее время используются различные методы, учитывающие особенности каждой категории.

Таким образом, задача определения таких атрибутов личности, как «пол», «раса» и «возраст» является актуальной, а модификация существующих и разработка новых методов компьютерного зрения,

ориентированных на автоматическое решение этой задачи, приобретает все большую практическую ценность. При этом задача определения возраста наиболее востребована и является самой сложной, так как процесс старения проявляется у людей очень индивидуально, и даже человек плохо справляется с данной задачей.

1.3. Основные этапы определения атрибутов личности по изображению лица

Технологический процесс определения атрибутов личности по изображению лица содержит три этапа [61, 99, 101]:

- 1) нормализация изображения;
- 2) формирование вектора признаков, описывающего изображение;
- 3) классификация изображения.

Рассмотрим данные этапы.

В процессе нормализации изображение приводится к некоторому унифицированному виду, допускающему применение единых методов его обработки на последующих этапах.

Нормализация изображения предусматривает:

- распознавание лица и его выделение из общего изображения;
- преобразование изображения из цветного в одноканальное (оттенки серого);
- выделение на изображении ключевых точек (как правило — глаз); выравнивание изображения относительно линии уровня глаз;
- обрезание изображения исходя из среднестатистических пропорций лица человека на основе расстояния между зрачками;
- понижение уровня шума в исходном изображении за счет проведения медианной фильтрации [30, 35], позволяющей убрать

резкие выбросы, что весьма актуально для реальных фотографий, по сравнению с фото, полученными в идеальных условиях;

- получение фронтального изображения лица из реального за счет применения аффинного преобразования или деформации на основе триангуляции Делоне [112].

Перечисленные процедуры, выполняемые в процессе нормализации изображения, являются достаточно сложными и могут быть предметом отдельного исследования, выходящего за пределы настоящей работы. В настоящем анализе предполагается, что этап нормализации уже выполнен, и исследуются последующие этапы технологического процесса — формирование вектора признаков изображения и его классификация.

Для определения атрибутов личности по нормализованному изображению лица с помощью математических методов необходимо выполнить формальное описание исходного изображения. Далее этим описанием можно оперировать, используя какой-либо математический аппарат. Таким описанием является вектор признаков (числовых характеристик) изображения, каждый элемент которого несет информацию об определенном свойстве объекта [7]. На этапе формирования вектора признаков изображения (в англоязычной литературе используется термин – *features extraction*), осуществляется формирование пространства признаков и определение численных значений выбранных признаков, т.е. на этом этапе применяется некоторый метод, производящий замену реального изображения на набор численных значений характеризующих его признаков.

При решении задачи классификации следует выделить важный случай признакового описания объектов, существенной чертой которых является наличие структуры (химические соединения, кристаллические решетки, графы). Особенности построения обобщенных понятий на

основе анализа объектов, обладающих структурой, обсуждаются в ряде работ [31, 32, 34].

Использование информации о структуре объекта может предоставить дополнительные возможности для классификации. Данный подход с успехом используется для решения задачи классификации и распознавания текстовых документов [4, 5, 17, 18, 33]. Однако подобных примеров для решения задачи определения атрибутов личности по изображению лица в литературе не обнаружено.

Задача классификации заключается в разбиении множества изображений, заданных их векторами признаков, на классы (метки классов) по определенному критерию [7, 23, 24] (в нашем случае по полу, расе и возрасту). Объекты в пределах одного класса считаются эквивалентными с точки зрения критерия разбиения. Этап классификации приводит к конечной цели - определению искомым атрибутов личности по выбранным на предыдущем этапе признакам, т.е. происходит отнесение анализируемого изображения (его вектора признаков) каким-либо методом классификации к соответствующему классу (например: мужчина, монголоид, 27 лет).

Необходимо отметить, что проведенный в процессе исследования анализ литературных источников, посвященных рассматриваемой тематике, не выявил корреляции между методом формирования вектора признаков изображения и методом классификации. Это означает, что исследователь, решая задачу определения атрибутов личности по изображению лица, волен выбрать наиболее подходящий с его точки зрения метод формирования вектора признаков, а затем независимо от этого выбрать метод классификации. Так, например, в работах [62, 104] использован одинаковый метод формирования вектора признаков изображения при разных методах классификации, а в работах [87, 105, 112] наоборот: после использования разных методов формирования

вектора признаков изображения, на этапе классификации был использован один и тот же подход. Поэтому ниже проводится независимый анализ существующих методов формирования вектора признаков и методов классификации изображений лица с целью выявления наиболее перспективных для дальнейшего развития при определении искомых атрибутов личности.

1.4. Задача классификации изображений как задача обучения по прецедентам

Задача автоматической классификации изображений лиц людей по значениям атрибутов личности («пол», «раса» и «возраст») является типичной задачей обучения по прецедентам [8, 11, 13]. Неформально данная задача может быть сформулирована следующим образом: имеется множество \mathbf{I} изображений лиц людей (*объектов*), каждое из которых объективно характеризуется значением (или набором значений) определенного атрибута личности из множества значений (или наборов значений) \mathbf{Y} атрибута. То есть, объективно существует функция $\varphi: \mathbf{I} \rightarrow \mathbf{Y}$. Имеется подмножество \mathbf{I}' , $\mathbf{I}' \subset \mathbf{I}$, для элементов i которого известны значения $\varphi(i)$. Соответствующие пары $(i, \varphi(i))$, $i \in \mathbf{I}'$ называются *прецедентами* и рассматриваются как *выборка* \tilde{G}' , $\tilde{G}' \subset \tilde{G}$ (из графика \tilde{G} функции φ).

На основе множества прецедентов, т.е. выборки, необходимо сформировать алгоритм, который для всякого нового *объекта* i , $i \notin \mathbf{I}'$, сформирует *ответ* $\varphi^*(i)$, и, более того, оценить качество алгоритма. Такой алгоритм принято называть *решающей функцией*, а способ построения этой функции по множеству прецедентов — *методом обучения* [11, 13].

Как видим, решающая функция $\varphi^*(i)$ аппроксимирует функцию φ . При этом множество прецедентов \tilde{G}^l , $\tilde{G}^l \subset \tilde{G}'$, использующихся в

процессе формирования *решающей функции*, называют *обучающей выборкой*. Для оценки способности решающей функции решать указанную задачу, т.е. для оценки качества аппроксимации, а тем самым и качества прогнозируемых этой функцией решений, используется *тестовая выборка* – множество $\tilde{G}^t, \tilde{G}^t \subset \tilde{G}'$, прецедентов, которых нет в выборке $\tilde{G}^l: \tilde{G}^l \cap \tilde{G}^t = \emptyset$. Обучающая и тестовая выборки определены на множествах $\mathbf{I}^l = \text{Pr}_1(\tilde{G}^l)$ и $\mathbf{I}^t = \text{Pr}_1(\tilde{G}^t)$, где $\text{Pr}_1(\cdot)$ – операция проектирования бинарного отношения по первой координате.

Решающей функции φ^* соответствует функционал качества $\tilde{Q}(\varphi^* \tilde{G}^l, \tilde{G}^t)$ и функционал потерь $\tilde{L}(\varphi^* \tilde{G}^l, \tilde{G}^t)$, оценивающие по определенным правилам результат применения функции φ^* , полученный по обучающей выборке \tilde{G}^l и тестовой выборке \tilde{G}^t .



Рис. 1.1. Пример прецедента «объект – ответ»

Данную задачу можно сформулировать следующим образом.

Множество \mathbf{I} нормализованных изображений заменяется множеством \mathbf{X} , $\mathbf{X} \in \mathbb{R}^m$ описывающих эти изображения векторов

признаков (именно вектор признаков изображения рассматривается далее как *объект*), m – размерность пространства признаков. Это означает, что применяется некоторый метод формирования вектора признаков изображения $\Psi: \mathbf{I} \rightarrow \mathbf{X}$ из некоторого множества Ψ таких методов. Будем обозначать $\Psi(\mathbf{I})$ множество объектов, соответствующих подмножеству изображений \mathbf{I}' , $\mathbf{I}' \subseteq \mathbf{I}$, полагая, что на множестве \mathbf{I}' отображение $\Psi': \mathbf{I}' \rightarrow \Psi(\mathbf{I}')$ является биекцией. Для данного множества *объектов* $\mathbf{X} = \Psi(\mathbf{I})$ объективно имеется множество *ответов* \mathbf{Y} , т.е. объективно существует функция $f: \mathbf{X} \rightarrow \mathbf{Y}$ (соответствует функции $\varphi: f(\Psi(i)) = \varphi(i)$). Её график обозначим G . При этом выборке \tilde{G}' соответствует выборка G' из графика G функции f . Элементы этой выборки получаются из прецедентов $(i, \varphi(i))$ выборки \tilde{G}' как $(\Psi(i), \varphi(i))$. При этом из обучающей \tilde{G}^l и тестовой \tilde{G}^t выборок получаются формализованные обучающая G^l и тестовая G^t выборки.

Задачей машинного обучения является нахождение аппроксимации функции f – функции $f_\mu: \mathbf{X} \rightarrow \mathbf{Y}$, параметры которой определяются по обучающей выборке G^l , т.е. множеству пар $(x_i, y_i) \in G^l$, $x_i = \Psi(i) \in \mathbf{X}^l$, $i \in \mathbf{I}^l$, $y_i = \varphi(i)$, $i = 1, \dots, n$.

Способ формирования решающей функции называют *методом обучения* $\mu: \mu \in \mathbf{M}$, где \mathbf{M} – множество известных методов обучения.

Пусть $\mu(G^l) = f_\mu \in \mathbf{F}$ – некоторая *решающая функция*, сформированная с помощью метода обучения $\mu \in \mathbf{M}$ с использованием формализованной обучающей выборки G^l . Тогда значение $f_\mu(x)$ называют предсказанием решающей функции f_μ на объекте x .

Функционал качества предсказаний полученной по формализованной обучающей выборке G^l решающей функции f_μ определяется по формализованной тестовой выборке G^t , которая состоит из формализованных описаний прецедентов – пар (x_i, y_i) , $x_i \in \mathbf{X}^t$, $y_i = \varphi(i)$, i

$= 1, \dots, k$, которые при обучении не использовались, и обозначается $Q(f_\mu, G^l, G^t)$. Его надо определить так, чтобы он служил оценкой обобщающей способности решающей функции f_μ , т.е. способности данной функции предсказывать (с известной долей или иной мерой ошибок) ответ y для объекта $x \in \mathbf{X}^t$ и $x \notin \mathbf{X}^l$. Иногда вместо функционала качества используется функционал потерь $L(f_\mu, G^l, G^t)$.

Относительно рассмотренной выше неформализованной постановки задачи функционалу качества $\tilde{Q}(\Psi, \varphi^*, \tilde{G}^l, \tilde{G}^t)$ и функционалу потерь $\tilde{L}(\Psi, \varphi^*, \tilde{G}^l, \tilde{G}^t)$ соответствует функционал качества $Q_\Psi(f_\mu, G^l, G^t)$ и функционал потерь $L_\Psi(f_\mu, G^l, G^t)$, определяемые для различных способов Ψ формализации изображения и метода μ обучения.

В зависимости от множества *ответов* \mathbf{Y} различают следующие задачи классификации [11, 13]:

- задача бинарной классификации, если $|\mathbf{Y}|=2$; например, определение атрибута «пол»: $\mathbf{Y} \in \{\text{«мужчина»}, \text{«женщина»}\}$; при бинарной классификации обычно полагают, что $\mathbf{Y} = \{-1, +1\}$;
- задача множественной классификации, если $|\mathbf{Y}|>2$; например, определение атрибута «раса»: $\mathbf{Y} = \{\text{«европеоидная»}, \text{«монголоидная»}, \text{«негроидная»}\} = \{-1, 0, 1\}$, или возрастной группы: $\mathbf{Y} \in \mathbb{R}^l$, где l – число возрастных групп, а \mathbf{Y} – например, их нижние границы.
- задача восстановления регрессии, если $\mathbf{Y} \in \mathbb{R}^m$; например, определение атрибута «возраст»: $\mathbf{Y} = [5, 100]$.

Задачи множественной классификации иногда решаются на основе разложения в задачи бинарной классификации, а задача восстановления формулируется с учетом результата решения задачи множественной классификации, например, определяется возрастная группа и для неё прогнозируется значение возраста лица, отнесенного к ней на этапе

определения возрастной группы. Тогда решающие функции задачи представляются как суперпозиции функций, соответствующих подзадачам.

Таким образом задача определения атрибутов личности по нормализованному изображению лица как задача обучения с прецедентами из обучающей \tilde{G}^l и тестовой \tilde{G}^t выборок формулируется следующим образом: определить метод $\psi \in \Psi$ формирования множества X векторов признаков нормализованных изображений из множества I , метод обучения $\mu \in M$ и решающую функцию $f_\mu \in F$, аппроксимирующую неизвестную зависимость $f: \Psi(I) \rightarrow Y$, и такую, что определённый функционал качества $Q_\psi(f_\mu, G^l, G^t)$ имеет максимальное значение или (в двойственной формулировке) функционал потерь $L_\psi(f_\mu, G^l, G^t)$ имеет минимальное значение.

Решение этой оптимизационной задачи получается сравнением решений, из конечного множества частных оптимизационных задач, соответствующих различным методам формирования векторов признаков изображений и различным методам представления решающих функций. Например, для каждого метода формализации определяется наилучший метод обучения $\mu(\psi)$:

$$\mu(\psi) = \operatorname{argmax}_{\Psi, \mu, Q_\psi}(f_\mu, G^l, G^t) \quad (1.1)$$

или

$$\mu(\psi) = \operatorname{argmin}_{\Psi, \mu, L_\psi}(f_\mu, G^l, G^t) \quad (1.1')$$

Затем выбирается наилучший метод формализации

$$\psi = \operatorname{argmax}_{\Psi, Q_\psi}(f_\mu, \Psi) \quad (1.2)$$

или

$$\psi = \operatorname{argmin}_{\Psi, L_\psi}(f_\mu, \Psi) \quad (1.2')$$

Таким образом, для каждого выбора соответствующей пары методов возникает классическая задача оптимизации – выбор параметров решающей функции данного метода. Поэтому в решении оптимизационной задачи в целом различают переборный этап со сравнением наилучших решений, получаемых классическими методами для каждого варианта выбора. Отметим, что рассмотренный подход к постановке и решению оптимизационной задачи поиска наилучших методов формализации и обучения не учитывает затраты времени при формализации, построении решающей функции и ее применении для классификации. Поэтому наилучшие по такому выбору методы могут оказаться неприемлемым по производительности.

Поскольку все известные методы представления решающих функций предполагают, что эти функции определяются на векторных представлениях объектов из множества X , прежде всего, мы рассмотрим методы формирования таких описаний. Затем проанализируем известные методы оценки качества решений для задач с множествами ответов указанных трех групп и, наконец, опишем известные методы решения соответствующих частных оптимизационных задач построения решающих функций при бинарной и множественной классификации и в задачах регрессии.

1.5. Анализ методов формирования вектора признаков изображения лица для определения атрибутов личности

При использовании методов компьютерного зрения для определения атрибутов личности по изображению лица в процессе признакового описания нормализованных изображений I' применяются различные

методы Ψ формирования вектора признаков X' . Эти методы можно разделить на две большие группы [61]:

- группа методов Ψ_t , основанная на использовании текстуры (значений интенсивности пикселей) изображения;
- группа методов Ψ_a , основанная на выделении антропометрических точек на изображении лица человека с последующим выделением информации о расстоянии между этими точками и их взаимном расположении.

К группе Ψ_t относится множество методов [40, 62, 70, 93, 104, 105], в которых используются либо непосредственно значения интенсивности пикселей нормализованных изображений, либо результаты применения некоторых операторов к значениям интенсивностей. Результирующее признаковое описание представляет собой гистограмму значений интенсивностей или их модификаций. Отметим, что прямое использование пикселей изображения для решения задачи классификации возможно только для изображений небольшой размерности и в задачах определения атрибутов личности практически не используется, а находит применение в задаче выделения лиц на общем изображении [39].

В последние годы наибольшее распространение в рамках группы методов Ψ_t получил метод *локальных бинарных шаблонов* (*Local Binary Patterns, LBP*) [70, 84, 88, 105, 110], что объясняется целым рядом важных особенностей получаемого с его помощью вектора признаков изображения. Локальный бинарный шаблон – это определенный вид признака, представляющий собой описание окрестности пикселя изображения в двоичном представлении. Базовый оператор *LBP*, применяемый к пикселю изображения, использует восемь пикселей окрестности и принимает значение интенсивности центрального пикселя в качестве порога. Пиксели со значением интенсивности большим или равным значению интенсивности центрального пикселя принимают

значения равные «1», остальные принимают значения равные «0». Таким образом, результатом применения базового оператора **LBP** к пикселю изображения является восьмиразрядный бинарный код, который описывает окрестность этого пикселя (запись производится по часовой стрелке, начиная с верхнего левого пикселя). Далее этот код рассматривается как двоичная запись некоторого числа, и именно это число и ставится в соответствие данному пикселю (рис. 1.2).

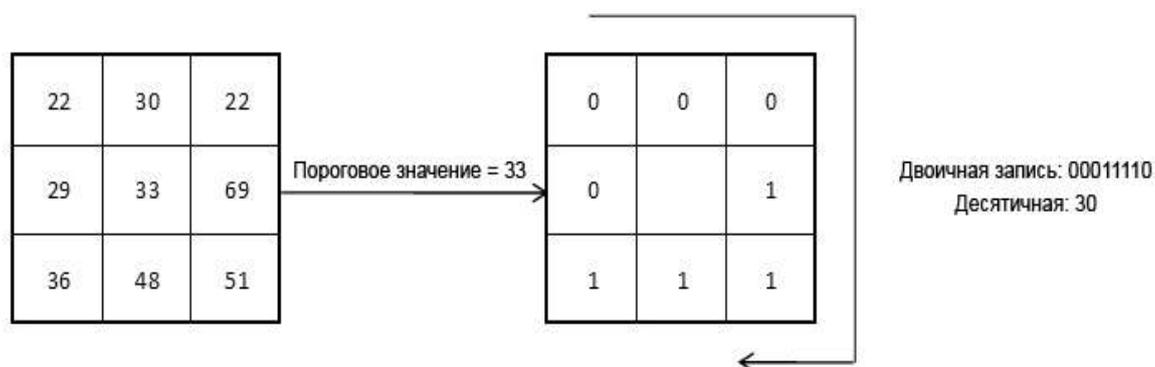


Рис. 1.2. Пример применения оператора **LBP**

После получения для каждого пикселя значения **LBP** строится гистограмма этих значений, т.е. формируется набор

$$\Psi_{LBP}(i) = \bar{x}_i = (\bar{0}_i, \bar{1}_i, \dots, \bar{x}_i, \dots, \bar{255}_i)$$

частот \bar{x}_i появления бинарных шаблонов, имеющих десятичные числовые эквиваленты x , $x \in \{0, 1, \dots, 255\}$, в изображении i : $\bar{x}_i \in \mathbb{R}$, $0 \leq \bar{x}_i < 1$.

При этом коллекция \mathbf{I}' из m элементов обучающей и k элементов тестовой выборки представляется матрицами размером $m \times 256$ и $k \times 256$:

$$\Psi_{LBP}(\mathbf{I}') = \mathbf{X}' = \begin{pmatrix} \bar{0}_1 & \bar{1}_1 & \dots & \bar{x}_1 & \dots & \bar{255}_1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \bar{0}_i & \bar{1}_i & \dots & \bar{x}_i & \dots & \bar{255}_i \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \bar{0}_m & \bar{1}_m & \dots & \bar{x}_m & \dots & \bar{255}_m \end{pmatrix}$$

$$\Psi_{LBP}(\mathbf{I}^t) = \mathbf{X}^t = \begin{pmatrix} \bar{0}_1 & \bar{1}_1 & \dots & \bar{x}_1 & \dots & \bar{255}_1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \bar{0}_i & \bar{1}_i & \dots & \bar{x}_i & \dots & \bar{255}_i \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \bar{0}_k & \bar{1}_k & \dots & \bar{x}_k & \dots & \bar{255}_k \end{pmatrix}$$

Соответствующими векторами заменяются элементы i в описаниях прецедентов (i, y_i) в обучающей \tilde{G}^l и тестовой \tilde{G}^t выборках.

Применительно к изображениям лиц локальные бинарные шаблоны характеризуют *локальные особенности изображения*, а частоты конкретных локальных особенностей можно рассматривать как некоторую обобщенную модель изображения лица и принимать описанное выше отображение $\Psi_{LBP}(i)$ в качестве метода формирования векторной модели изображения лица и матричной модели представления коллекции таких изображений.

Используемые при формировании $\Psi_{LBP}(i)$ частоты \bar{x}_i появления бинарных шаблонов относятся к методу tf-взвешивания [116]. Примеров использование других методов взвешивания (tf-idf, tfc, ltc [116]) при формировании модели изображения на основе **LBP**, в доступной литературе не обнаружено.

Важно отметить, что гистограмма, построенная для всего изображения в целом, кодирует лишь наличие тех или иных локальных особенностей, но при этом не содержит никакой информации об их расположении на изображении. Для учета такого рода информации нормализованное изображение разбивается на s регионов, для каждого из которых вычисляется своя гистограмма **LBP**, т.е. для региона j изображения i будет сформирована гистограмма

$$\Psi_{LBP}(i_j) = \bar{\mathbf{x}}_{i_j} = (\bar{0}_{i_j}, \bar{1}_{i_j}, \dots, \bar{x}_{i_j}, \dots, \bar{255}_{i_j}).$$

Результирующим описанием изображения i является вектор значений \bar{x}_i , сформированный как конкатенация s гистограмм, полученных по s регионам исходного изображения (см. Рис. 1.3):

$$\Psi_{LBP}(i) = \bar{x}_i = (\bar{0}_{i_1}, \bar{1}_{i_1}, \dots, \bar{x}_{i_1}, \dots, \bar{255}_{i_1}, \dots, \bar{0}_{i_j}, \bar{1}_{i_j}, \dots, \bar{x}_{i_j}, \dots, \bar{255}_{i_j}, \dots, \bar{0}_{i_k}, \bar{1}_{i_k}, \dots, \bar{x}_{i_k}, \dots, \bar{255}_{i_k})$$

Понятно, что данное описание изображения уже не будет являться гистограммой, так как сумма значений его элементов равна s .

Исходя из соотношений геометрических размеров характерных элементов лица человека (глаз, носа, рта, бровей и пр.) нормализованное изображение традиционно разбивается на $k = 6 \times 7 = 42$ региона [105].

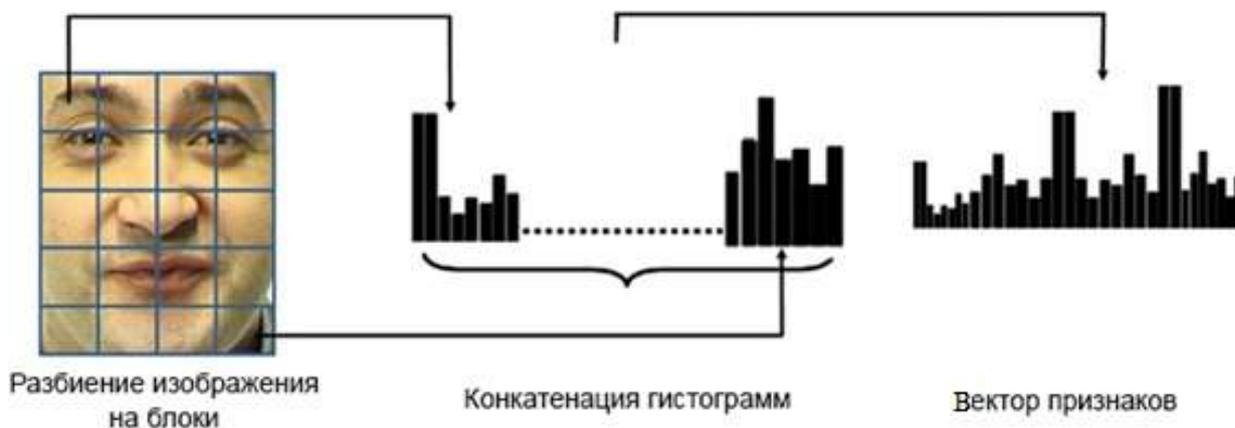


Рис. 1.3. Пример формирования вектора признаков изображения, разделенного на $4 \times 5 = 20$ регионов, методом **LBP** [88]

Более крупное разбиение приводит к потере информации о расположении локальных особенностей, а более мелкое - существенно увеличивает размерность пространства признаков без повышения качества описания, что усложняет процесс последующей классификации.

Благодаря особенностям применения оператора **LBP**, локальные бинарные шаблоны инвариантны к небольшим изменениям в условиях освещенности и небольшим поворотам классифицируемого изображения [96], что и обусловило их широкое распространение для решения задач определения таких атрибутов личности, как «пол» [37, 49, 84, 92], «раса»

[42, 68, 111] и возрастная группа [47, 64, 103]. Примеров использования метода **LBP** для определения численного значения атрибута «возраст» методами восстановления регрессии в доступной литературе не обнаружено. Основной причиной этого является очень большая размерность пространства признаков, формируемого данным методом. Так при разбиении изображения на $6 \times 7 = 42$ блока размерность пространства признаков $m = |\bar{x}_i| = 42 \times 256 = 10752$.

В работе [96] показано, что не все бинарные шаблоны обладают одинаковой информативностью. Наиболее важными являются *равномерные* шаблоны, к которым относятся бинарные комбинации, содержащие не более трех серий «0» и «1» (например, 00000000, 001110000 и 11100001). Равномерные **LBP** определяют только важные локальные особенности изображения: концы линий, грани, углы и пятна (см. Рис. 1.4).

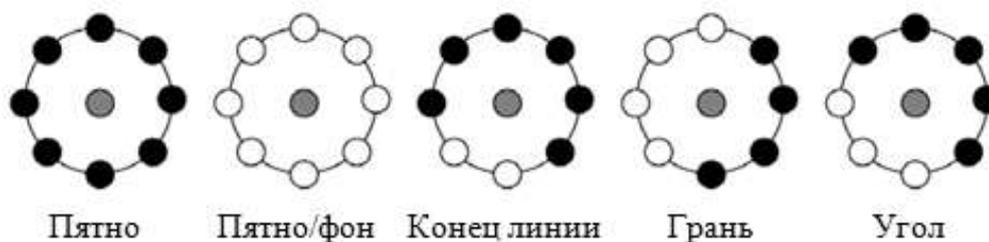


Рис. 1.4. Примеры равномерных шаблонов **LBP**, описывающих различные локальные особенности изображения

Использование только равномерных **LBP** приводит к существенному сокращению размерности пространства признаков, так как в этом случае используется только $p(p-1)+2=58$ шаблонов [96]). В этом случае при разбиении изображения на 42 блока размерность $|\bar{x}_i|$ пространства признаков $r = 42 \times 58 = 2436$, что почти в 4,5 раза меньше исходного, но все равно остается достаточно большой.

Итогом применения оператора **LBP** при использовании разбиения изображения на 42 блока и использовании *равномерных* шаблонов будет следующий вектор признаков для изображения i :

$$\Psi_{LBP_{\text{равн.}}}(i) = \bar{x}_i = (\bar{0}_{i_1}, \bar{1}_{i_1}, \bar{2}_{i_1, \dots}, \bar{x}_{i_1, \dots}, \bar{57}_{i_1, \dots}, \bar{0}_{i_j, \dots}, \bar{x}_{i_j, \dots}, \bar{57}_{i_j, \dots}, \bar{0}_{i_{42}, \dots}, \bar{x}_{i_{42}, \dots}, \bar{57}_{i_{42}, \dots}),$$

где \bar{x} есть десятичный числовой эквивалент равномерного шаблона с порядковым номером x при упорядоченности равномерных шаблонов по их числовым эквивалентам, а \bar{x}_{i_j} – частота встречаемости такого шаблона среди вхождений равномерных шаблонов в полном векторном представлении j -го фрагмента i -го изображения.

Важным достоинством данного метода является простота реализации оператора **LBP**, что позволяет использовать его в задачах обработки изображений в реальном времени.

Методологически близкий к **LBP** метод – построение гистограммы направленных градиентов (*Histogram of Oriented Gradients* - **HOG** [60, 102]) основан на вычислении градиента изменения интенсивности для каждого пикселя изображения и формирования гистограммы данных градиентов для различных участков изображения (при этом выделяется восемь базовых направлений: «север», «северо-восток», «восток» и т.д.). В итоге формируется пространство признаков близкое по своим характеристикам к методу **LBP**, но его размерность плохо поддается сокращению, а процедура вычисления градиента интенсивности заметно сложнее процедуры формирования **LBP**. Основное применение данный метод формирования пространства признаков изображения и аналогичный ему метод формирования гистограмм направления края изображения (*Edge orientation histogram* – **ЕОН** [83]) нашел при решении задачи выделения заданного объекта на изображении, а не при определении атрибутов личности. Каких либо преимуществ от

использования методов *HOG/EON* по сравнению с использованием метода *LBP* при формировании вектора признаков изображения для дальнейшей классификации в доступной литературе не обнаружено.

Базовая идея методов второй группы – Ψ_a заключается в воспроизведении процесса описания лица, характерном для человека: фиксируются особые точки и их расположение на лице (границы глаз, носа, рта, бровей, подбородка и пр.). Среди модификаций методов, относящихся к этой группе [47, 51, 64, 74, 103], наибольшее распространение для определения атрибутов личности по изображению лица получил метод, основанный на формировании активной модели внешности (*Active Appearance Models*, *AAM*, [51]). Рассмотрим идею данного метода, выделив его достоинства и недостатки.

Данная модель является развитием активной модели формы (*Active Shape Models*, *ASM*, [52]) путем добавления в нее модели текстуры. Первым этапом данного метода является построение активной модели формы (*ASM*) путем разметки нормализованного изображения лица - выделения на нем некоторых пронумерованных характерных антропометрических точек. Качественное описание изображения достигается при числе таких точек n порядка 60 - 70. Пример такой разметки приведен на Рис. 1.5 [118]

Автоматическое определение важных антропометрических точек основано на выделении характерных особенностей изображения: Т-образные пересечения краев, углы, легко различимые биологические особенности, границы изображения и пр. [52].

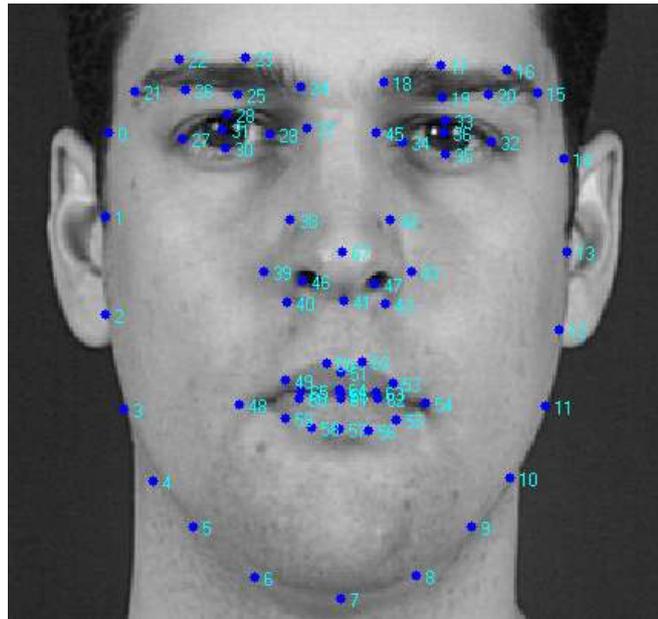


Рис. 1.5. Пример разметки изображения лица антропометрическими точками

Главной предпосылкой для построения *ASM* является наблюдение, что между расположением антропометрических точек есть зависимости. Для моделирования этих зависимостей по обучающей выборке \tilde{G}^l строится статистическая модель положения важных антропометрических точек на изображении лица. При этом в качестве признакового описания изображения используются координаты выделенных точек на нормализованных изображениях, т.е. формируется набор

$$\Psi_{ASM}(i) = \bar{\mathbf{x}}_i = (x_1, x_2, \dots, x_u, y_1, y_2, \dots, y_u)^T,$$

а вся совокупность из n прецедентов обучающей выборки \tilde{G}^l образует матрицу

$$\Psi_{ASM}(\tilde{G}^l) = \mathbf{X} = (\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_i, \dots, \bar{\mathbf{x}}_n).$$

Пример изображения антропометрических точек, соответствующий матрице \mathbf{X} , приведен на Рис. 1.6 [118]. При построении модели предполагается, что «облака» точек (выделены эллипсами) подчиняются нормальному распределению [52].

После выделения главных компонент [11] указанной матрицы получаем следующее выражение для синтезированной формы:

$$\mathbf{X} = \bar{\mathbf{X}} + \mathbf{P}_{sh} \mathbf{B}_{sh}$$

$$\mathbf{B}_{sh} = \mathbf{P}_{sh}^T (\mathbf{X} - \bar{\mathbf{X}})$$

Здесь $\bar{\mathbf{X}}$ — форма изображения, усредненная по всем реализациям обучающей выборки (базовая форма), \mathbf{P}_{sh} — матрица собственных векторов, \mathbf{B}_{sh} — вектор параметров формы.

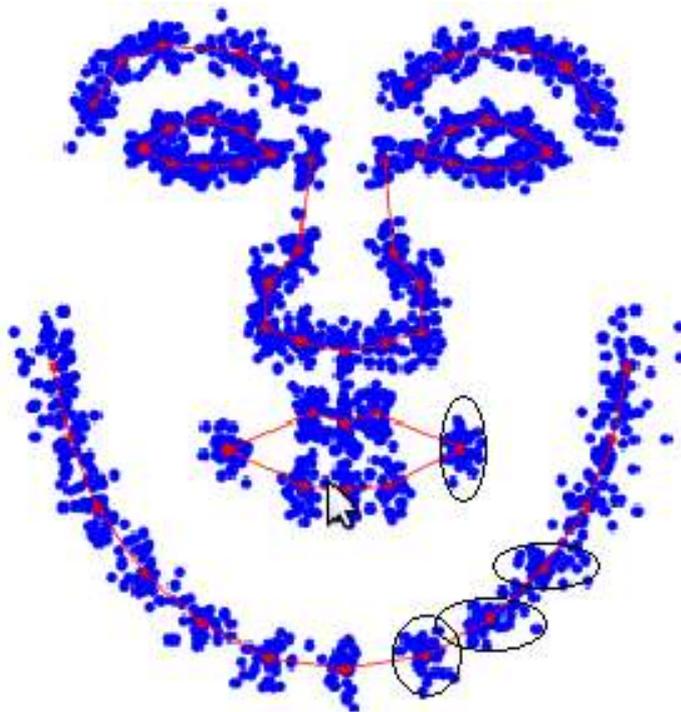


Рис. 1.6. Пример расположения антропометрических точек, сформированных по обучающей выборке

Приведенное выражение означает, что форма любого изображения \mathbf{X} может быть выражена как сумма базовой формы $\bar{\mathbf{X}}$ и линейной комбинации собственных форм, содержащихся в матрице \mathbf{P}_{sh} . Меняя

значения \mathbf{B}_{sh} можно синтезировать различные формы изображений (см. Рис. 1.7 [118]).

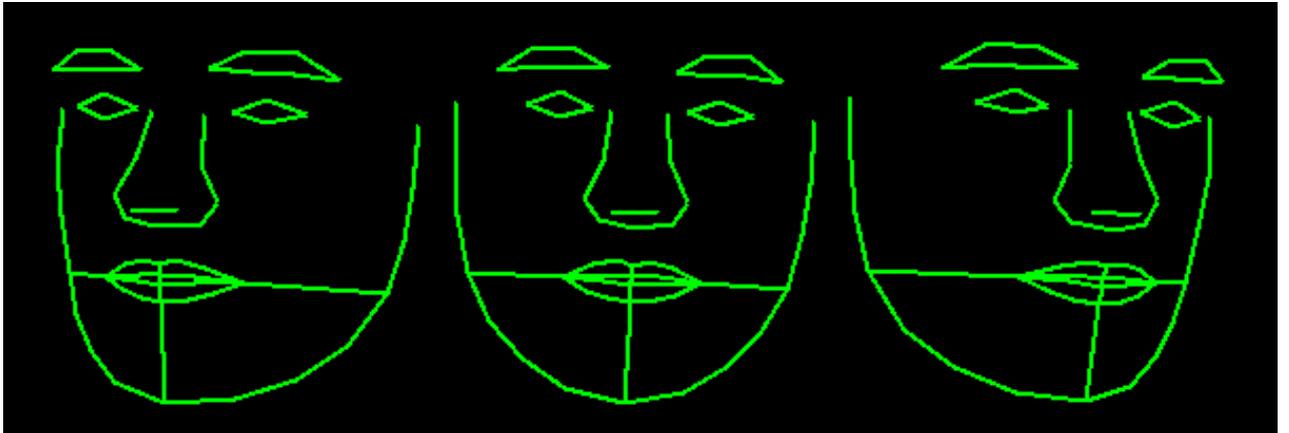


Рис. 1.7. Пример влияния значения параметров \mathbf{B}_{sh} на форму изображения

На втором этапе построения модели \mathbf{AAM} учитывается информация о текстуре (значениях интенсивности пикселей) изображения. Для этого изображение разбивается на непересекающиеся треугольники, вершинами которых являются выделенные антропометрические точки активной модели формы. Пример разбиения изображения на треугольники для формирования модели текстуры приведен на Рис. 1.8 [118]

Далее для каждого треугольника вычисляется среднее значение интенсивности, т.е. изображение описывается вектором

$$\Psi_{\mathbf{AAM}}(i) = \bar{\mathbf{z}}_i = (z_1, z_2, \dots, z_\nu)^T,$$

где ν – количество треугольников в изображении, а вся совокупность из n прецедентов обучающей выборки $\tilde{\mathbf{G}}^l$ образует матрицу

$$\Psi_{\mathbf{AAM}}(\tilde{\mathbf{G}}^l) = \mathbf{Z} = (\bar{\mathbf{z}}_1, \dots, \bar{\mathbf{z}}_i, \dots, \bar{\mathbf{z}}_n).$$

Аналогично построению активной модели формы по методу главных компонент, строится приближенная модель текстуры

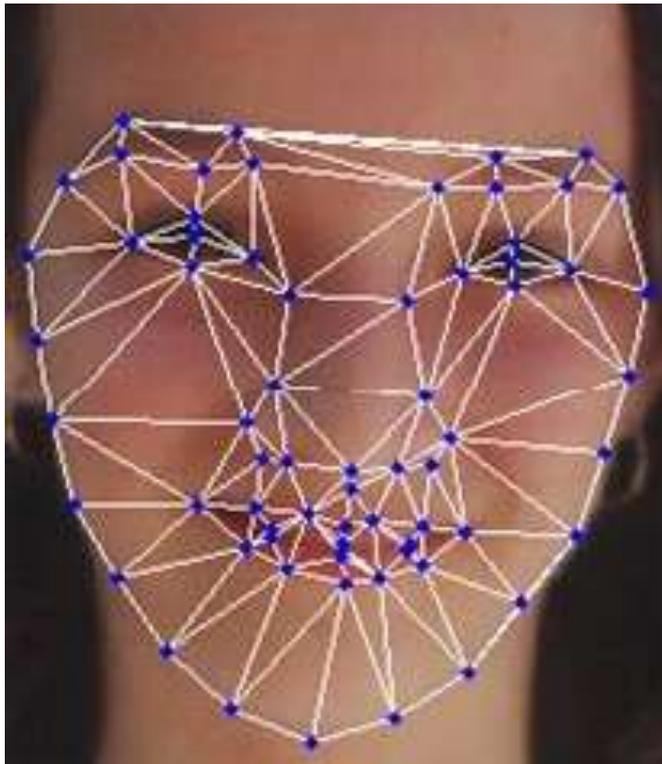


Рис. 1.8. Пример разбиения изображения для построения модели текстуры в рамках ААМ

изображения:

$$\mathbf{Z} = \bar{\mathbf{Z}} + \mathbf{P}_t \mathbf{B}_t$$

$$\mathbf{B}_t = \mathbf{P}_t^T (\mathbf{Z} - \bar{\mathbf{Z}}).$$

Как и в модели формы, текстура любого изображения \mathbf{Z} может быть синтезирована суммой базовой текстуры $\bar{\mathbf{Z}}$ и линейной комбинации собственных форм, содержащихся в матрице \mathbf{P}_t . Меняя значения \mathbf{B}_t можно синтезировать различные текстуры изображений.

Общее признаковое описание изображения i по методу ААМ представляет собой конкатенацию векторов $\bar{\mathbf{x}}_i$ и $\bar{\mathbf{z}}_i$.

С помощью данного метода формирования вектора признаков для изображений лиц сегодня получены хорошие результаты при определении таких атрибутов личности, как «пол» и «раса» [72, 82, 103]. Это связано с тем, что именно взаимное расположение важных антропометрических

точек, совместно с текстурой изображения, позволяют определить принадлежность личности к определенному полу и расе. Что касается атрибута «возраст», то данный подход формирования вектора признаков используется для определения возрастной группы [47, 64, 103], а не для вычисления прогнозируемого значения возраста на основе восстановления регрессии. Это можно объяснить тем, что значимые изменения в расположении антропометрических точек и расстояния между ними в первую очередь зависят от пола и расы личности, и практически не меняются при изменении возраста на несколько лет для людей в диапазоне от 20 до 50 лет [50].

Самой большой проблемой данного метода формирования вектора признаков является сложность автоматического выделения на изображении характерных антропометрических точек. При этом для выделения указанных точек изображение все равно анализируется на уровне текстуры, т.е. интенсивности его пикселей, так как другой информации об изображении просто нет. Следовательно, процесс формирования активной модели внешности включает в себя и процесс анализа пикселей изображения, и процесс построения модели формы.

Отметим, что метод *ААМ* в наибольшей степени ориентирован на решение другой задачи из области компьютерного зрения - идентификации личности по изображению лица, так как в этом случае разметку эталонного изображения осуществляет человек, интуитивно «разумно» выделяя наиболее важные и характерные антропометрические точки. Построенная «вручную» модель внешности затем успешно используется для поиска требуемой личности на других изображениях [56] или определения выражения эмоций [38].

Основываясь на проведенном анализе методов признакового описания изображений, для формирования пространства и значений признаков в настоящей работе признано наиболее перспективным

использование метода локальных бинарных шаблонов *LBP*. Основными причинами такого вывода являются:

- простота применения оператора *LBP* по сравнению с выделением антропометрических точек, что сокращает затраты времени на формирование вектора признаков изображения;
- наличие в результирующем признаке изображения информации не только о наличии локальных особенностей изображения, но и об их расположении, что сближает два эти подхода;
- возможность использования только тех шаблонов, которые хранят больше информации о локальных особенностях изображения;
- устойчивость результата применения оператора *LBP* к изменению освещенности и небольшим поворотам изображения;
- наличие принципиальных сложностей в определении значения атрибута «возраст» с помощью восстановления регрессии при использовании методов, основанных на выделении и анализе антропометрических точек, из-за слабого изменения в расположении антропометрических точек с возрастом для людей среднего возраста.

Главной проблемой на пути успешного применения метода *LBP* для формирования вектора признаков изображения является большая размерность результирующего пространства (более 2000). Поэтому важным направлением совершенствования данного метода является снижение этой размерности. В настоящей работе для этого предлагается использовать априорную информацию о человеческом лице для выделения наиболее значимых признаков при формировании признакового описания изображения (данный вопрос будет рассмотрен в главе 2).

1.6. Оценка качества методов определения атрибутов личности

Для выбранных способа $\Psi \in \Psi$ векторного представления изображения и метода $\mu \in \mathbf{M}$ формирования решающей функции f_μ решается задача оптимизации значений ее параметров для достижения наилучшего решения по прогнозированию значения того или иного атрибута личности. Сравнением решений этих частных оптимизационных задач при различных вариантах выбора методов Ψ и μ достигается решение общей задачи (1.1 – 1.2) или (1.1' – 1.2').

Все рассматриваемые ниже показатели качества $Q_{\Psi, \mu}$ и показатели потерь $L_{\Psi, \mu}$ решающих функций f_μ являются общеизвестными [**Ошибка! источник ссылки не найден.**, 91, 108], и задача настоящего параграфа – выделить те показатели качества, которые нашли практическое применение именно при решении задачи классификации изображений лиц людей по исследуемым атрибутам («пол», «раса», «возраст»).

Следует различать оценку показателей качества при бинарной (или множественной) классификации, и при решении задачи восстановления регрессии, так как в первом случае (например, определение атрибутов «пол» или «раса») можно оценить лишь долю правильно или неправильно классифицированных объектов, а во втором случае (прогнозирование значения атрибута «возраст») – не только долю, но и значение величины ошибки.

В случае бинарной и множественной классификаций очевидным и, потому, часто используемым, является функционал качества $Q_{\Psi}(f_{\mu}, \mu(\psi), G^l, G^t)$, который определяет долю верно классифицированных построенной по формализованной методом ψ обучающей выборке G^l решающей функцией $f_{\mu(\psi)}$ объектов, представленных в прецедентах тестовой выборки G^t [**Ошибка! Источник ссылки не найден.**]. Данная

ера качества называется *Accuracy* (точность) и определяется следующим образом:

$$Q_{\Psi}(f_{\mu(\psi)}, G^l, G^t) = k_{correct}/k, \quad (1.3)$$

где $k_{correct}$ - количество правильно классифицированных объектов, а k - размер тестовой выборки.

При этом можно определить и функционал потерь:

$$L_{\Psi}(f_{\mu(\psi)}, G^l, G^t) = 1 - Q_{\Psi}(f_{\mu(\psi)}, G^l, G^t) = 1 - k_{correct}/k.$$

Ниже решающую функцию $f_{\mu(\psi)}$ будем сокращенно обозначать f , полагая способы формализации и построения решающей функции известными по умолчанию.

Важным условием применимости данного функционала качества Q (или потерь L) является примерное равенство количества объектов, относящихся к разным классам, в тестовой выборке. Легко видеть, что при существенном перекосе в количестве объектов в пользу какого-то одного класса в тестовой выборке, данный функционал может показывать очень хорошие результаты для решающей функции, которая вообще не способна выделять объекты другого класса. Поэтому, для оценки качества решающих функций при решении задач бинарной и множественной классификации в случае существенной неравномерности распределения объектов по классам используют метрики точности (*Precision*, **Pr**) и полноты (*Recall*, **Re**) [91]. Иногда они используются сами по себе, иногда в качестве базиса для производных метрик. Применительно к рассматриваемым в данной работе формализованным функциям f традиционно используемые понятия [91] описываются следующим образом: $t\{P(a,b)\}$ принимает целочисленное значение 1, если предикат $P(a,b)$ имеет значение true, и значение 0 в противном случае.

Традиционным является использование следующих понятий и обозначений [91]:

$$TP(f, X^l, X^t, y) = \sum_{x \subset X^t} t\{(f(x) = y) \& x \subset y\}$$

— число истинно-положительных решений, т.е. число объектов $x \in X^t$, принадлежащих классу y , и отнесенных решающей функцией f к классу y ;

$$TN(f, X^l, X^t, y) = \sum_{x \subset X^t} t\{(f(x) \neq y) \& x \not\subset y\}$$

— число истинно-отрицательных решений; т.е. число объектов $x \in X^t$, не принадлежащих классу y , и не отнесенных решающей функцией f к классу y ;

$$FP(f, X^l, X^t, y) = \sum_{x \subset X^t} t\{(f(x) = y) \& x \not\subset y\}$$

— число ложно-положительных решений; т.е. число объектов $x \in X^t$, не принадлежащих к классу y , но отнесенных решающей функцией f к классу y (ошибка первого рода);

$$FN(f, X^l, X^t, y) = \sum_{x \subset X^t} t\{(f(x) \neq y) \& x \subset y\}$$

— число ложно-отрицательных решений; т.е. число объектов $x \in X^t$, принадлежащих к классу y , но не отнесенных решающей функцией f к классу y (ошибка второго рода).

Тогда меры точности $Pr(f, X^l, X^t, y)$ и полноты $Re(f, X^l, X^t, y)$ определяются следующим образом [91]:

$$Pr(f, X^l, X^t, y) = \frac{TP(f, X^l, X^t, y)}{TP(f, X^l, X^t, y) + FP(f, X^l, X^t, y)};$$

$$Re(f, X^l, X^t, y) = \frac{TP(f, X^l, X^t, y)}{TP(f, X^l, X^t, y) + FN(f, X^l, X^t, y)}.$$

Смысл этих значений очевиден.

Полнота $Re(f, X^l, X^t, y)$ определяет долю верных предсказаний решающей функции f для объектов класса y среди всех объектов, отнесенных данной функцией к классу y (например, доля изображений мужчин среди всех изображений, отнесенных решающей функцией f к классу «мужчины»).

Точность $Pr(\varphi^*, X^l, X^t, y)$ определяет долю верных предсказаний решающей функции f для объектов класса y среди всех объектов, действительно относящихся к классу y (например, доля изображений мужчин, среди всех изображений мужчин, входящих в тестовую выборку).

Понятно, что чем выше точность и полнота, тем лучше. Поскольку максимальная точность и полнота не достижимы одновременно, в задачах классификации используют меру Ван Ризбергена [108] (F_β - мера), определяемую соотношением:

$$F_\beta = \frac{(1 + \beta^2) * Pr * Re}{\beta^2 * Pr + Re},$$

где β - определяет вес (важность) точности в сравнении с полнотой: при $0 < \beta < 1$ приоритет отдается точности, а при $\beta > 1$ - полноте. При оценке качества решающих функций в большинстве случаев используют значение $\beta = 1$, при котором метрика является сбалансированной (F_1 - мера), а ее значение равно среднему гармоническому точности и полноты [91, 108]:

$$F_1 = \frac{2 * Pr * Re}{Pr + Re}. \quad (1.3)$$

Отметим, что данная мера является функционалом качества, т.е. чем больше значение F_1 - меры, тем лучше решающая функция.

Таким образом, для оценки качества решающих функций при бинарной классификации (определение атрибутов «пол» и «раса»)

целесообразно использовать функционал качества *Accuracy* при примерном равенстве количества объектов разных классов в тестовой выборке \mathbf{X}^t и F_1 - меру — при наличии перекоса в количестве объектов в пользу какого-то одного класса.

При множественной классификации на основе нескольких бинарных классификаторов возникает необходимость вычислить единственный агрегированный показатель, объединяющий показатели отдельных классификаторов. Для этого существует два метода. При макроусреднении (*macroaveraging*) вычисляется обычное среднее значение по классам. При микроусреднении (*microaveraging*) значения точности и полноты вычисляются по суммарным показателям всех классов [17, 117].

Для задачи восстановления регрессии (определение значения атрибута «возраст») используются следующие две традиционные и вполне очевидные меры [48, 63]:

- Средняя абсолютная ошибка (*Mean Absolute Error, MAE*):

$$MAE(f, G^t) = \frac{1}{k} \sum_{i=1}^k |y_i - f(\bar{x}_i)|. \quad (1.4)$$

Данная мера характеризует среднее абсолютное отклонение предсказанных решающей функцией f значений искомого атрибута для объектов $\bar{x}_i \in \mathbf{X}^t$ от значений этого атрибута в прецедентах тестовой выборки для данного объекта; k – по-прежнему количество объектов в тестовой выборке. При определении атрибута «возраст» по тестовой выборке данная мера покажет среднюю ошибку (в количестве лет), допускаемую решающей функцией. Отметим, данная метрика является функцией потерь, т.е. чем меньше значение $MAE(f, G^t)$, тем лучше решающая функция f .

- Суммарный счёт (*Cumulative Score, CS(ε)*); $\text{sign}(x)=1$, если $x > 0$; $\text{sign}(x)=0$, если $x = 0$;

$$CS(f, G^t, \varepsilon) = \frac{1}{k} \sum_{i=1}^k \text{sign}(\varepsilon - |y_i - f(\bar{x}_i)|). \quad (1.5)$$

Здесь $(-.)$ – усеченное вычитание, т.е. 0 вместо отрицательного результата.

Данная мера характеризует долю правильных предсказаний решающей функцией f значений искомого атрибута для объекта $\bar{x}_i \in \mathbf{X}^t$, где правильно предсказанными (относящимися к **TP**) считаются те объекты, для которых величина ошибки e в предсказания $f(\bar{x}_i)$ возраст относительно значения y_i в соответствующем прецеденте тестовой выборки не превышает заданного значения ε ; k - количество объектов в тестовой выборке G^t . При определении атрибута «возраст» по тестовой выборке при заданном уровне ошибки (например, 3 года) данная мера покажет долю объектов (изображений людей), классифицированных решающей функцией с ошибкой в определении возраста менее 3 лет. Отметим, что данная мера является функционалом качества, т.е. чем больше значение $CS(f, G^t, \varepsilon)$, тем лучше решающая функция f .

Две последние рассмотренные меры широко используются для оценки качества решающих функций, предсказывающих значение атрибута «возраст», и будут использованы в настоящей работе.

1.7. Анализ методов классификаций изображений лиц людей по различным атрибутам

1.7.1. Методы классификации изображений по атрибуту «пол»

Общие положения.

Классификация изображений лиц людей с целью определения атрибута «пол» является типичной задачей бинарной классификации, т.е. решающая функция f принимает значения:

$$f(x) = \begin{cases} +1, & \text{«мужчина»;} \\ -1, & \text{«женщина»}. \end{cases}$$

Данная задача является одной из самых распространенных среди задач определения атрибутов личности. Первые работы по ее решению восходят еще к 90-м годам прошлого века [67]. Практический интерес к этой задаче в последующие годы породил большое количество работ, в которых предлагались различные подходы (различные методы обучения μ):

- многоуровневые нейронные сети [36, 39, 67, 106];
- линейный дискриминантный анализ (*LDA*) в различных модификациях [44, 57, 86];
- бустинг в различных модификациях [40, 70, 105, 111];
- метод опорных векторов (*SVM*) [105, 84, 93, 76, 54, 103, **Ошибка! сточник ссылки не найден.**].

Необходимо отметить, что при использовании перечисленных методов классификации, для формирования вектора признаков \mathbf{X} изображения применялись как методы группы Ψ_a , основанные на выделении антропометрических точек [62, 93, 103], так и методы группы Ψ_t , использующие для этой цели текстуру изображения [54, 84, 105] (см. параграф 1.5).

Важным этапом на пути развития методов классификации изображений по атрибуту «пол» стала работа [92], опубликованная в 2008 году, в которой авторы провели сравнение лучших на тот момент методов классификации изображений по указанному атрибуту. В указанной работе были проанализированы следующие методы:

- многоуровневая нейронная сеть;
- метод опорных векторов – *Support Vector Mashine*, *SVM* (формирование вектора признаков \mathbf{X} как на основе активной модели

внешности - *AAM*, так и на основе локальных бинарных шаблонов - *LBP*);

- метод бустинга – *Adaboost* - адаптивное улучшение (в трех модификациях: *Threshold Adaboost* (пороговое адаптивное улучшение), *LUT Adaboost* (табличное адаптивное улучшение), *Mean Adaboost* (адаптивное улучшение в среднем)).

Отметим, что сравнение методов проводилось как на специально подготовленной и размеченной открытой базе изображений лиц (FERET, [98]), так и на изображениях лиц, полученных непосредственно из сети Интернет. При этом в качестве критерия для сравнения использовалась метрика *Accuracy* (см. параграф 1.6), определяющая долю верно классифицированных объектов. Поскольку в тестовой и обучающей выборках количество представителей разных классов (мужчин и женщин) было примерно одинаковым, использование этой метрики было естественным и обоснованным. Ниже приведен фрагмент одной из итоговых таблиц, представленных в данной работе, полученной авторами при классификации изображений лиц из сети Интернет (см. Табл. 1.1).

Таблица 1.1.

Сравнение точности методов классификации по атрибуту «пол» на изображениях из сети Интернет [98]

№ п/п	Метод классификации	<i>Accuracy</i> , %
1	Многоуровневая нейронная сеть	65,96
2	Метод опорных векторов совместно с <i>AAM</i>	66,48
3	Метод опорных векторов совместно с <i>LBP</i>	67,25
4	Метод бустинга 1 (<i>Threshold Adaboost</i>)	66,29
5	Метод бустинга 2 (<i>LUT Adaboost</i>)	66,19
6	Метод бустинга 3 (<i>Mean Adaboost</i>)	66,14

Сравнение на реальных изображениях из сети Интернет представляется практически более значимым, чем сравнение на базе специально подготовленных изображений FERET. Ожидаемо для всех сравниваемых методов точность классификации по метрике *Accuracy* на изображениях, полученных из сети Интернет, была ниже (примерно, на 15%), чем на изображениях базы FERET [49] (см. Табл. 1.2).

Таблица 1.2.

Сравнение точности методов классификации по атрибуту «пол» на изображениях из базы FERET [49]

№ п/п	Метод классификации	<i>Accuracy</i> , %
1	Метод опорных векторов (HI kernel) совместно с <i>LBP</i>	88,42
2	Метод опорных векторов (LI kernel) совместно с <i>LBP</i>	84,93

Анализ представленных результатов позволяет сделать два вывода.

1) Проанализированные методы классификации изображений лиц по атрибуту «пол» показывают примерно одинаковые результаты (различие в точности определения пола по метрике *Accuracy* не превосходят всего 1.5%). Это означает, что в процессе развития этих методов классификации, в итоге все они привели к практически одинаковому уровню качества, и любой из них может быть использован для дальнейшего совершенствования. Тем не менее, лидером, хотя и с небольшим отрывом, является метод опорных векторов (*SVM*) в сочетании с формированием вектора признаков методом локальных бинарных шаблонов (*LBP*). Учитывая ранее отмеченный факт простоты применения оператора *LBP* (см. параграф 1.5), данный результат является еще одним объяснением, почему именно метод *LBP* находит все большее распространение при формировании вектора признаков изображения в задачах определения атрибутов личности.

2) Существующие на сегодня методы классификации изображений лиц по атрибуту «пол» позволяют правильно классифицировать примерно две трети реальных изображений, что подтверждает актуальность задачи их модернизации с целью повышения качества классификации.

Результаты аналогичного сравнения методов классификации по атрибуту «пол» [49], опубликованные позже (в 2011 году), подтвердили преимущество использования метода *SVM* в сочетании с *LBP* над другими подходами (достигнута точность классификации более 90%, но на специально подготовленной базе данных изображений лиц людей, что несколько снижает значимость данного результата).

Анализ доступных работ, выполненных по тематике определения атрибута «пол» по изображению лица за последующие годы, показал, что во всех этих работах для решения рассматриваемой задачи используют либо метод опорных векторов *SVM* [41, 55, 75, 107], либо метод бустинга в различных модификациях [100, 105]. Поэтому идеи именно этих методов предлагается использовать в настоящей работе для дальнейшего развития.

Прежде, чем подробно рассмотреть указанные методы, необходимо отметить, что в последнее время получила дополнительное развитие идея использования многослойных нейронных сетей большой размерности для решения задач идентификации и классификации изображений [45, 46, 95]. С помощью алгоритмов «глубокого обучения» (*Deep Learning*) получены выдающиеся результаты по идентификации и классификации объектов. Главной проблемой практического внедрения подобных систем является чрезвычайно высокие требования к производительности аппаратной платформы, способной решать указанные задачи (задача обучения такой нейронной сети относится к категории НРС - *High Performance Computing*). В подобных системах используется десятки тысяч процессорных ядер (как общего назначения, так и GPU), а время обучения измеряется многими сутками [95]. Поэтому, подобные решения

технологически и экономически пока не могут конкурировать с другими подходами при решении задачи определения атрибутов личности в системах компьютерного зрения. Сферой применения алгоритмов «глубокого обучения» на сегодня остается, например, диагностика в медицине, где подобные системы зачастую успешно конкурируют с человеком [95].

Метод опорных векторов (*Support Vector Machine* - *SVM*) был разработан в 60-е годы коллективом советских математиков под руководством В.Н.Вапника [8, 53]. Рассмотрим его применительно к задаче бинарной классификации [11, 13].

Пусть имеется обучающая выборка G^l , $|G^l| = n$, заданная множеством пар прецедентов (\bar{x}_i, y_i) , $i = 1, n$; $\bar{x}_i \in \mathbb{R}^m$, $y_i \in \{-1, +1\}$. Множество F_{SVM} , из которого выбираются решающие функции по методу опорных векторов, образовано функциями вида

$$f(\bar{x}) = \text{sign}(\langle \bar{w}, \bar{x} \rangle + w_0), \quad (1.6)$$

где \langle, \rangle — скалярное произведение векторов [1], \bar{w} — ортонормированный вектор к разделяющей классы гиперплоскости, w_0 — вспомогательный параметр (сдвиг гиперплоскости). Такие функции и образуют множество F_{SVM} .

Те объекты, для которых $f(\bar{x}) = +1$ попадают в один класс, а объекты, для которых $f(\bar{x}) = -1$ — в другой. Выбор именно такой функции неслучаен, так как любая гиперплоскость может быть задана в виде уравнения

$$\langle \bar{w}, \bar{x} \rangle + w_0 = 0$$

для некоторых \bar{w} и w_0 .

Базовая идея метода *SVM* заключается в нахождении таких \bar{w} и w_0 , которые максимизируют расстояние между классами, так как максимизация ширины разделяющей классы полосы (отступа) приводит к более уверенной классификации объектов [11, 13] (см. Рис. 1.7).

При оптимальном расположении разделяющей гиперплоскости все граничные объекты, принадлежащие разным классам, должны находиться на одинаковом расстоянии от этой плоскости. Тогда условие

$$-1 < \langle \bar{w}, \bar{x} \rangle + w_0 < +1$$

задает полосу, разделяющую классы. Важно, что ни одна из точек множества \mathbf{X}^l не может лежать внутри полосы, а границами полосы являются две параллельные гиперплоскости, проходящие через точки (объекты), ближайšie к разделяющей гиперплоскости.

При этом сама разделяющая гиперплоскость проходит ровно посередине полосы. Те объекты \bar{x}_i из прецедентов (\bar{x}_i, y_i) обучающей выборки G^l (или объекты \bar{x}_i множества \mathbf{X}^l), которые лежат на граничных гиперплоскостях, т.е. удовлетворяют соотношению

$$y_i = \langle \bar{w}, \bar{x}_i \rangle + w_0,$$

называются *опорными* векторами (*Support Vectors*).

В этом случае для всех $\bar{x}_i \in \mathbf{X}^l$ справедливо соотношение:

$$\langle \bar{w}, \bar{x}_i \rangle + w_0 = \begin{cases} \leq -1, & \text{если } y_i = -1; \\ \geq +1, & \text{если } y_i = +1. \end{cases}$$

Данное соотношение эквивалентно следующему:

$$y_i \cdot (\langle \bar{w}, \bar{x}_i \rangle + w_0) \geq 1$$

В работе [13] показано, что ширина разделяющей классы полосы равна $2/\|\bar{w}\|$, т.е. ширина полосы максимальна, когда норма вектора [1] \bar{w} минимальна.

Проблема нахождения максимума расстояния сводится к задаче нахождения минимума $\|\bar{w}\|^2$, которая является стандартной задачей квадратичного программирования [13]

$$\begin{cases} \|\bar{w}\|^2 \rightarrow \min \\ y_i \langle \bar{w}, \bar{x}_i \rangle + w_0 \geq 1 \end{cases} \quad (1.7)$$

и решается методом множителей Лагранжа.

Если $\lambda = (\lambda_1, \lambda_2 \dots \lambda_n)$ - вектор двойственных переменных метода множителей Лагранжа, то $\lambda_i=0$ для всех объектов, не являющихся опорными векторами, так как они расположены на большем расстоянии от разделяющей гиперплоскости, чем опорные.

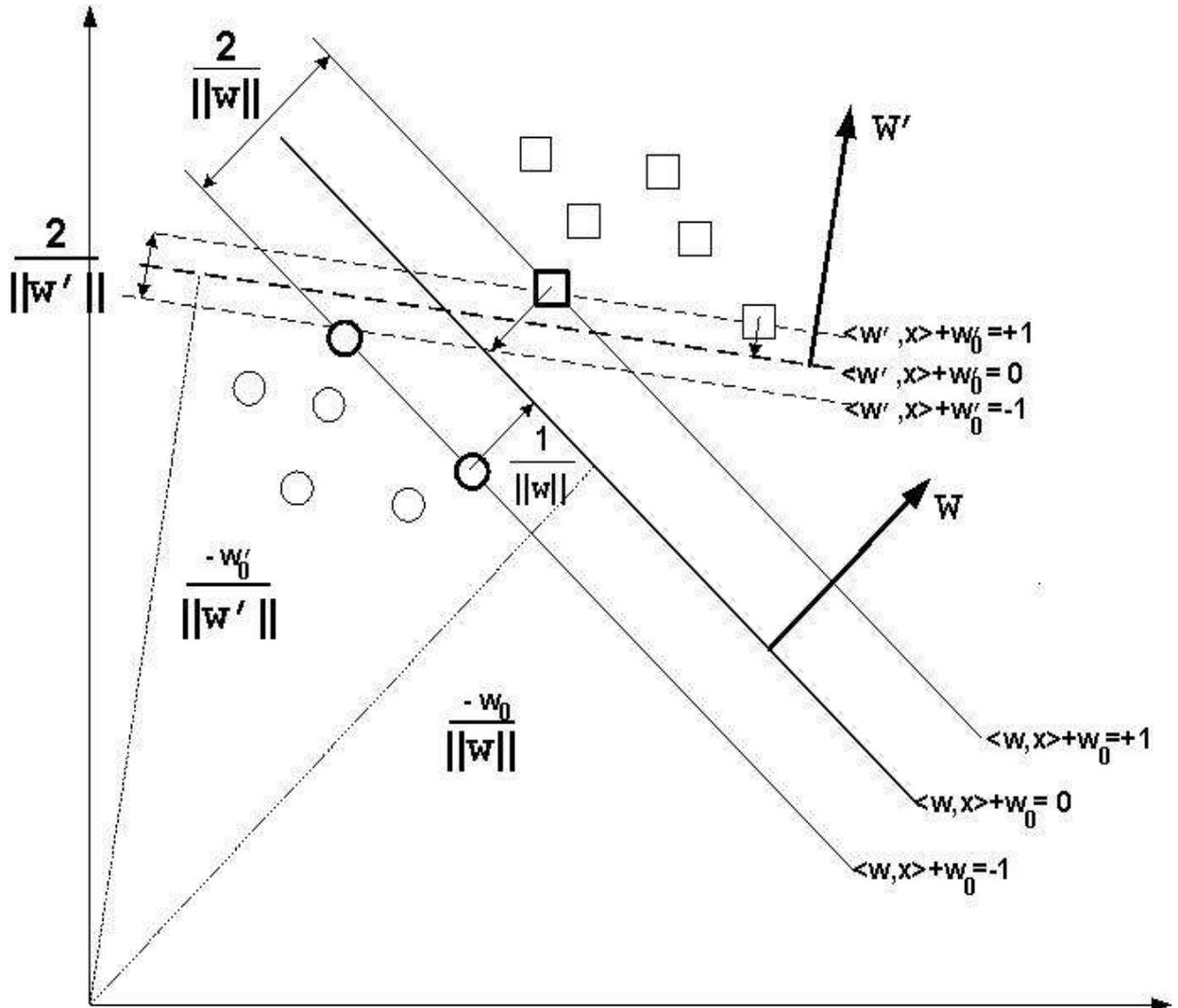


Рис. 1.9. Графическое пояснение идеи метода SVM

Задача квадратичного программирования, содержащая только двойственные переменные λ_i , формулируется следующим образом [11, 13]:

$$\begin{cases} -\sum_{i=1}^n \lambda_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j y_i y_j \langle \bar{x}_i, \bar{x}_j \rangle \rightarrow \min; \\ \sum_{i=1}^n \lambda_i y_i = 0; \\ \lambda_i \geq 0. \end{cases} \quad (1.8)$$

Стандартное решение данной задачи приводит к следующему итоговому соотношению для решающей функции [13]:

$$f(\bar{x}) = \text{sign}\left(\sum_{i=1}^n \lambda_i y_i \langle \bar{x}_i, \bar{x} \rangle + w_0\right), \quad (1.9)$$

где параметр w_0 рекомендуется вычислять как медиану по всем опорным векторам [11, 13], т.е.

$$w_0 = \text{med}\{\langle \bar{w}, \bar{x}_i \rangle - y_i\}.$$

Отметим, что в сумме (1.9) участвуют только опорные векторы, для которых $\lambda_i \neq 0$. Это означает, что другие объекты обучающей выборки не влияют на результат, т.е. они могут быть исключены. Указанная особенность метода **SVM** называется свойством *разреженности* [13] и позволяет построить компактный классификатор (решающую функцию).

Все приведенные рассуждения справедливы, если исходная обучающая выборка \mathbf{X}' является линейно разделимой, что практически не встречается. Для обобщения метода **SVM** на случай линейной неразделимости, решающей функции f разрешают допускать ошибки на обучающей выборке, но эти ошибки минимизируют. Если $\xi_i \geq 0$ - величина ошибки для объекта \bar{x}_i (шум), то вместо неравенства в (1.7) имеет место неравенство $y_i \langle \bar{w}, \bar{x}_i \rangle + w_0 \geq 1 - \xi_i$ [11, 13]:

$$\begin{cases} \frac{1}{2} \|\bar{w}\|^2 + C \sum_{i=1}^n \xi_i \rightarrow \min \\ y_i \langle \bar{w}, \bar{x}_i \rangle + w_0 \geq 1 - \xi_i. \end{cases} \quad (1.10)$$

Положительная константа C является управляющим параметром метода и позволяет находить компромисс между максимизацией ширины разделяющей полосы и минимизацией суммарной ошибки (см. Рис. 1.10)

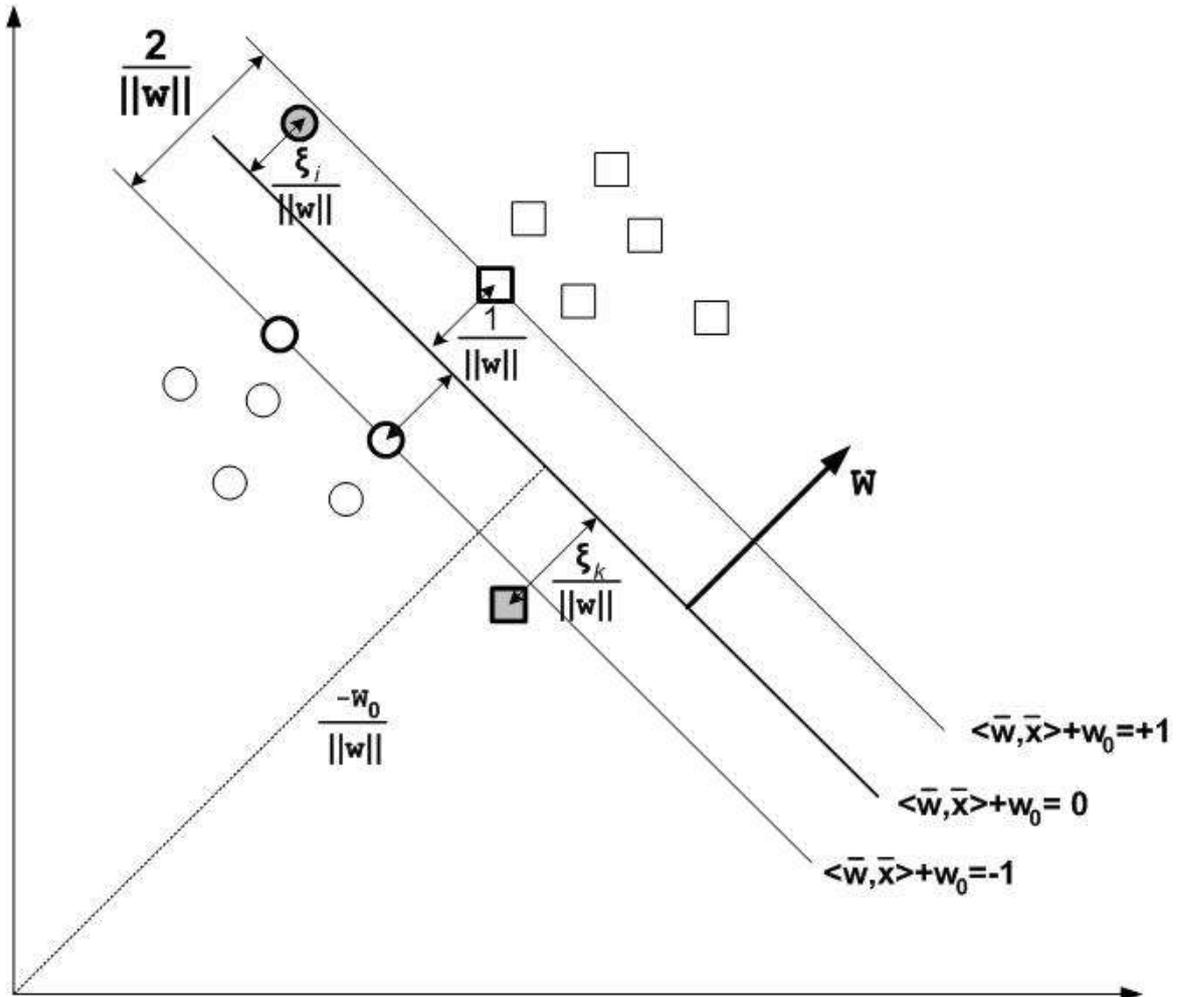


Рис. 1.10. Классификация объектов методом *SVM* при наличии «шума»

С учетом сказанного, задача квадратичного программирования (1.10) принимает вид [11, 13]:

$$\left\{ \begin{array}{l} -\sum_{i=1}^n \lambda_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j y_i y_j \langle \bar{x}_i, \bar{x}_j \rangle \rightarrow \min; \\ \sum_{i=1}^n \lambda_i y_i = 0; \\ 0 \leq \lambda_i \leq C. \end{array} \right. \quad (1.11)$$

Единственное отличие от линейно-разделимого случая заключается в том, что имеется ограничение сверху $0 \leq \lambda_i \leq C$. Данный вариант алгоритма называют *SVM* с «мягким» зазором (*soft-margin SVM*), тогда как в линейно-разделимом случае говорят об *SVM* с «жестким» зазором (*hard-margin SVM*) [13].

Если есть основания полагать, что обучающая выборка почти линейно-разделима, но содержит лишь некоторые объекты-выбросы, то применяют процедуру фильтрации выбросов. Сначала задача решается при некотором C , а затем из выборки удаляются объекты, имеющие наибольшее значение ξ_i , после чего задача решается заново на усеченной выборке.

Рассмотренный способ решения проблемы линейной неразделимости помогает при наличии некоторого «шума», препятствующего линейной разделимости объектов обучающей выборки \mathbf{X}^l . Наряду с этим, весьма распространен случай, когда линейная неразделимость объектов носит функциональный характер. Решением проблемы в этом случае [11, 13] является переход от исходного пространства вектора признаков \mathbf{X} к другому пространству большей размерности с помощью некоторого преобразования $\theta: \mathbf{X} \rightarrow \mathbf{H}$. Тогда в пространстве \mathbf{H} обучающая выборка \mathbf{X}^l становится линейно-разделимой, а пространство \mathbf{H} называют спрямляющим [13].

Если в методе *SVM* заменить признаки x_i на значения $\theta(x_i)$, то построение решающей функции методом *SVM* проводится так же, как и ранее. Важно, чтобы в новом пространстве существовало скалярное

произведение $\langle \theta(u), \theta(v) \rangle$. Функцию $K(u, v) = \langle \theta(u), \theta(v) \rangle$ называют *ядром*, если она симметрична и положительно определена [11, 13]. На Рис. 1.9 приведен пример использования ядра $K(u, v) = \langle u, v \rangle^2$, которому соответствует функция $\theta: \mathbb{R}^2 \rightarrow \mathbb{R}^3$.

При использовании спрямляющих пространств решающая функция f (1.9) по методу *SVM* принимает вид [13]:

$$f(\bar{x}) = \text{sign} \left(\sum_{i=1}^n \lambda_i y_i K(\bar{x}_i, \bar{x}) + w_0 \right).$$

При решении практических задач классификации изображений по атрибуту «пол» чаще всего используется RFB-ядро (*Radial Basis Function*) вида [11, 13]

$$K(\bar{x}_i, \bar{x}) = \exp(-\gamma \|\bar{x}_i - \bar{x}\|^2),$$

которое вычисляет оценку близости вектора (объекта) \bar{x} к опорному вектору \bar{x}_i , где γ - параметр.

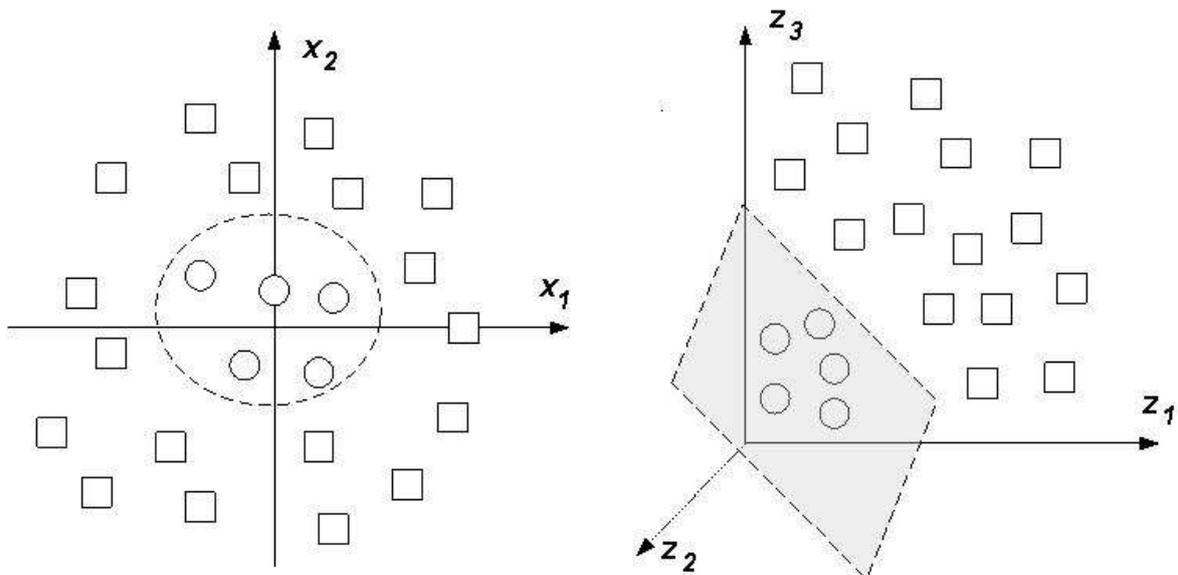


Рис. 1.11. Пример применения спрямляющего пространства:
 $(x_1, x_2) \rightarrow (z_1, z_2, z_3) := (x_1^2, \sqrt{2} x_1 x_2, x_2^2)$

Метод бустинга (*Adaboost*) [11, 13].

При решении сложных задач классификации часто оказывается, что ни одна из решающих функций не обеспечивает желаемого качества классификации. В таких случаях имеет смысл построить композицию решающих функций, в которой ошибки отдельных из них взаимно компенсируются.

Наиболее общее определение алгоритмической композиции дается в алгебраическом подходе Ю.И.Журавлева [20, 21]. Для задач бинарной классификации самым распространенным примером композиции являются различные способы взвешенного голосования [13], к которым относится и *Adaboost*.

Пусть имеется некоторый набор из m решающих функций бинарной классификации $b: b(x) = \{+1, -1\}$, которые принято называть базовыми алгоритмами [13]. Тогда искомая решающая функция f определяется следующим образом [13]:

$$f(\bar{x}) = \text{sign} \left(\sum_{i=1}^m \alpha_i b_i(\bar{x}) \right), \quad (1.12)$$

где $\alpha_i \in \mathbb{R}^m$ - положительные и нормированные веса базовых алгоритмов, т.е.

$$\sum_{i=1}^m \alpha_i = 1.$$

В данном случае композицию называют *выпуклой композицией* базовых алгоритмов [13].

Функционал потерь $L(f(\bar{x}), G^l)$ композиции определяют как число ошибок, допускаемых этой композицией на обучающей выборке G^l :

$$L(f(\bar{x}), G^l) = \sum_{j=1}^n \frac{1 - z_j}{2};$$

$$z_j = y_j \sum_{i=1}^n \alpha_i b_i(\bar{x}_j).$$

Для минимизации данного функционала предполагают [11, 13], что при добавлении в сумму очередного базового алгоритма b_m с его весом α_m оптимизируется только данный алгоритм, а все остальные полагаются фиксированными. Пороговая функция потерь $[z < 0]$ заменяется непрерывно дифференцируемой оценкой сверху, в качестве которой используют функцию $E(z) = \exp(-z)$ (выбор этой функции отличает базовый алгоритм *Adaboost* от его модификаций):

$$L(f(\bar{x}), G^l) \leq \tilde{L}(f(\bar{x}), G^l) = \frac{1}{2} \sum_{j=1}^n \left\{ 1 - \exp\left[-y_j \sum_{i=1}^m \alpha_i b_i(\bar{x}_j)\right] \right\}$$

Минимизация данного функционала приводит к следующему результату [11, 13]:

$$b_m(\bar{x}) = \arg \max_b \left(\sqrt{P(b_m)} - \sqrt{N(b_m)} \right);$$

$$\alpha_m = \frac{1}{2} \ln \frac{P(b_m)}{N(b_m)},$$

где $P(b_m)$ - суммарный вес правильных классификаций, а $N(b_m)$ - суммарный вес ошибочных классификаций базового алгоритма b_m ; $P(b_m) + N(b_m) = 1$. Полученные весовые коэффициенты нормируются.

Главным результатом метода *Adaboost* является следующее утверждение: если $P(b) > N(b)$, т.е. некоторый базовый алгоритм несколько лучше простого гадания, то указанный подход приводит к минимизации функционала потерь $L[f(\bar{x})]$. Требование, чтобы каждый

базовый алгоритм классифицировал объекты несколько лучше простого гадания, является достаточно слабым и на практике почти всегда выполняется. Этого оказывается достаточно, чтобы гарантировать сходимость алгоритма *Adaboost* за конечное число шагов [13].

На Рис. 1.12 приведена графическая иллюстрация метода бустинга.

Многочисленные модификации метода *Adaboost*, используемые для классификации изображений лиц людей по атрибуту «пол» [40, 70, 92, 100, 111, 105] базируются на различных подходах к выбору базовых алгоритмов и различных способах вычисления весовых коэффициентов α_i . Основная цель новых модификаций - снижение трудоемкости алгоритма. Классическими базовыми алгоритмами являются решающие деревья ограниченной глубины, начиная с одноуровневых решающих деревьев (*Decision Stump*, [71]).

Сравнение методов *SVM* и бустинга (*Adaboost*) позволяет сделать следующий вывод: использование метода опорных векторов является существенно менее трудоемким и, при этом, как показано в работах [49, 92,], дает большую точность классификации при определении атрибута «пол» на неподготовленных изображениях. Следовательно, именно этот метод целесообразно использовать для дальнейшего развития.

Однако идеи, заложенные в методе бустинга, могут оказаться полезными для дальнейшего рассмотрения. Поэтому, окончательный вывод о целесообразности использования того или иного подхода следует делать на основе анализа применимости рассмотренных методов для решения других задач (определения атрибута «раса» и атрибута «возраст»), поскольку в настоящей работе поставлена задача совместного определения нескольких атрибутов человека по изображению лица.

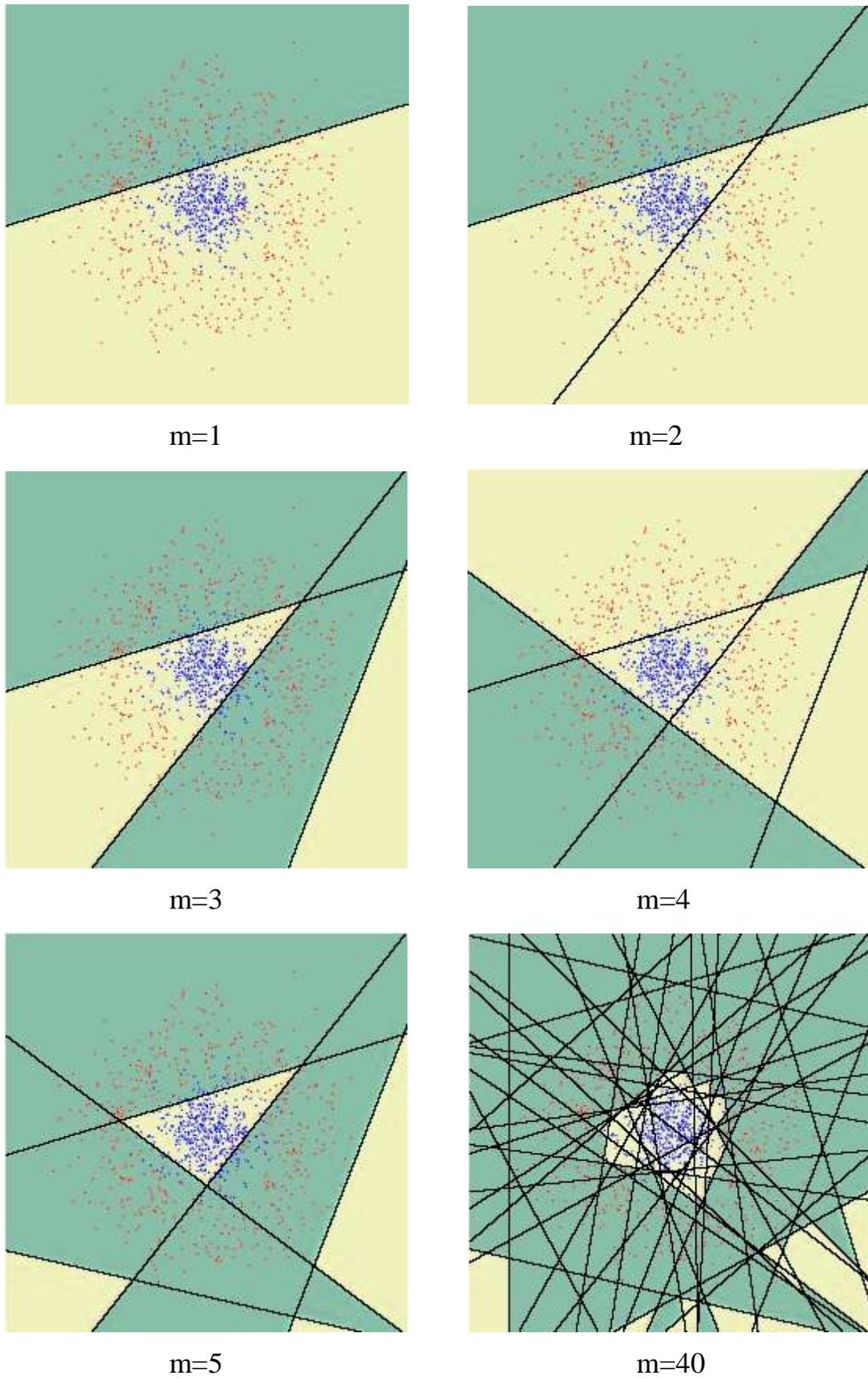


Рис. 1.12. Иллюстрация метода бустинга

1.7.2. Методы классификации изображений по атрибуту «раса»

Все человечество принято делить на три больших расы [6]: европеоидная (~40%), монголоидная (~30%), негроидная (негроидно-экваториальная) (~20%). Оставшиеся примерно 10% приходятся на смешанные расы. Поэтому, классификация изображений лиц людей по атрибуту «раса» является типичной задачей множественной классификации. Решению данной задачи посвящено значительное количество работ [54, 62, 67, 68, 72, 74, 86, 111], важной особенностью которых является сведение задачи множественной классификации к задаче бинарной классификации. Так, например, в работе [54] поочередно выделяются европеоидная раса и все остальные, негроидная раса и все остальные, а в работах [86, 111] - монголоидная раса и все остальные.

Подход, реализующий сведение задачи множественной классификации к последовательному применению бинарных классификаторов, является традиционным и получил название «один против всех» [7, 11, 13]. В рамках данного подхода строится бинарное дерево решающих функций $f \in \mathbf{F}$, каждая из которых выделяет только один класс объектов. Так в случае классификации по атрибуту «раса» на первом шаге объекты разделяются решающей функцией f_1 на два класса: «европеоиды» и все остальных. Если объект не попал в класс «европеоиды», то на втором шаге другая решающая функция f_2 производит разделение на класс «монголоиды» и остальные («негроиды»). Данный подход («один против всех») позволяет использовать огромное количество разработок в области бинарной классификации, часть которых (применительно к классификации по атрибуту «пол») была рассмотрена в предыдущем параграфе, для решения задач множественной классификации.

Анализ доступной литературы показал, что для классификации по атрибуту «раса» используются те же методы, что и для классификации по атрибуту «пол», а в последние годы наибольшее распространение, как и при классификации по атрибуту «пол», получил метод опорных векторов (*SVM*) [54, 62, 74]. При этом для формирования вектора признаков X используется как активная модель внешности *AAM* [62], так и локальные бинарные шаблоны *LBP* [74, 111]. Следует отметить, что метод бустинга также применяется для решения рассматриваемой задачи, хотя и с несколько худшим качеством [111].

Отметим, что указанное выше распределение человечества по расам естественным образом позволяет формировать тестовую выборку таким образом, что количество объектов в них находится в отношении примерно 2:1,5:1. Поэтому, для оценки предсказательной способности решающей функции при последовательном применении бинарной классификации (сначала отделяем «европеоидов», а затем из оставшихся выделяем «монголоидов») можно использовать критерий *Accuracy* (см. параграф 1.6).

Таким образом, анализ современного состояния методов классификации изображений по атрибуту «раса» не выявил каких-либо специфических подходов по сравнению с классификацией изображений по атрибуту «пол», т.е. для дальнейшего развития и для решения задачи классификации изображений лиц людей по атрибуту «раса» целесообразно использовать рассмотренные ранее методы

1.7.3. Методы классификации изображений по атрибуту «возраст»

Классификация изображений лиц людей по атрибуту «возраст» является наиболее сложной задачей. Как было указано в параграфе 1.4, данная задача относится к задаче восстановления регрессии, поскольку

решающая функция $f \in \mathbf{F}$, возвращает значение возраста, т.е. $f(x)=y$; $y \in [1,100]$.

Следует отметить, что в целом ряде работ задача классификации по атрибуту «возраст» сведена к задаче множественной классификации [47,, 48, 63, 64, 66, 102, 103] путем разделения всего возрастного диапазона на непересекающиеся классы (например, «дети», «подростки», «юноши/девушки» или конкретные диапазоны возрастов: (0 - 5), (6 - 12),(13 - 18),(19 - 25) и т.д.), которые могут пересекаться [**Ошибка! сточник ссылки не найден.**]. При этом для классификации используются те же методы, что и для классификации по атрибутам «пол» и «раса», рассмотренные ранее, т.е. задача сводится к последовательному применению бинарной классификации (алгоритм – «один против всех»). Вновь отметим, что в последние годы наиболее часто используется метод опорных векторов *SVM* [48, 64, 66, 103], хотя метод бустинга также находит применение [89, 105]. Такой подход существенно упрощает задачу классификации, но практически гораздо менее значим, чем определение конкретного значения возраста, так как существуют задачи (см. параграф 1.2), для которых определения только возрастной группы недостаточно. С другой стороны, определив значение возраста для некоторого объекта, не составляет труда отнести данный объект к какой-либо возрастной группе: если в процессе классификации определено, что возраст человека 40 лет, то отнесение его к классу «молодые» или к классу «старые» зависит только от цели классификации.

В работе [47] задача классификации по атрибуту «возраст» была сведена к последовательности бинарных классификаций: для каждого значения возраста был построен классификатор, который определял, старше человек на тестовом изображении, чем заданный возраст, или младше. Итоговое значение возраста определялось как число классификаторов, выдавших положительный результат. Отметим, что для

классификации вновь использовался метод *SVM*, а классификаторы различались только сдвигом гиперплоскости (см. параграф 1.7.1). В работе [103] данный подход получил дальнейшее развитие в направлении построения независимых классификаторов для каждого возраста. Наряду с этим, вместо минимизации значения ошибки было предложено минимизировать цену ошибки, т.е. некоторую функцию от значения ошибки. Использование данного подхода позволило получить на открытой базе изображений лиц людей MORPH [94] значение ошибки по критерию $MAE=6,07$ лет.

Поскольку определение значения возраста практически более востребовано, в настоящей работе задача классификации изображений лиц людей по атрибуту «возраст» рассматривается именно как задача восстановления регрессии [13]. Наибольшее распространение в последние годы получил метод регрессии на основе опорных векторов (*Support Vector Regression, SVR*, [61, 87, 109, 110]).

Рассмотрим формализованную обучающую выборку $G^l \subset G$, заданную множеством пар прецедентов $\{\bar{x}_i, y_i\}$, $i=1, n$; $G^l \in \mathbb{R}^m$, \bar{x}_i - вектор признаков, соответствующий изображению i обучающей выборки G^l , а y_i - соответствующее ему значение возраста; $Y \in \mathbb{R}$. Для многомерной линейной регрессии решающая функция принимает вид:

$$f(\bar{x}) = \langle \bar{w}, \bar{x} \rangle + w_0.$$

В качестве функции потерь используется кусочно-линейная ε -чувствительная функция $|z|_\varepsilon = \max\{0, |z| - \varepsilon\}$, которая не считает за ошибки отклонение $f(\bar{x})$ от y_i меньше ε [109]. Значение допустимой ошибки ε задает эксперт, исходя из постановки задачи классификации.

Для решения задачи вводятся две вспомогательные переменные ξ_i^+ и ξ_i^- :

$$\xi_i^+ = f(\bar{x}_i) - y_i - \varepsilon;$$

$$\xi_i^- = -f(\bar{x}_i) + y_i - \varepsilon.$$

Тогда задача сводится к минимизации функционала следующего вида [11, 13, 109]:

$$\begin{cases} \left\{ \frac{1}{2} \|\bar{w}\|^2 + C \sum_{i=1}^n (\xi_i^+ + \xi_i^-) \right\} \rightarrow \min_{w, w_0, \xi_i^+, \xi_i^-}; \\ y_i - \varepsilon_0 - \xi_i^- \leq (\langle \bar{w}, \bar{x}_i \rangle + w_0) \leq y_i + \varepsilon_0 + \xi_i^+; \\ \xi_i^+ \geq 0, \xi_i^- \geq 0, i = 1, n. \end{cases} \quad (1.13)$$

Данная задача сходна с задачей классификации по методу *SVM* (см. соотношение (1.10)) и также является задачей квадратичного программирования, которая решается методом множителей Лагранжа с использованием двойственных переменных λ_i^+ и λ_i^- .

Пропуская стандартные преобразования, запишем итоговое соотношение для решающей функции, возвращающей скалярное значение возраста для объекта \bar{x} , выраженное через двойственные переменные [11, 13]:

$$f(\bar{x}) = \sum_{i=1}^n (\lambda_i^+ + \lambda_i^-) \langle \bar{x}_i, \bar{x} \rangle + w_0$$

где параметр w_0 определяется из ограничений-неравенств, который становятся равенствами на опорных векторах:

$$f(\bar{x}_i) = \langle \bar{w}, \bar{x}_i \rangle + w_0 = \begin{cases} y_i + \varepsilon, & 0 < \lambda_i^+ < C, & \lambda_i^- = 0, \xi_i^+ = \xi_i^- = 0; \\ y_i - \varepsilon, & 0 < \lambda_i^- < C, & \lambda_i^+ = 0, \xi_i^+ = \xi_i^- = 0. \end{cases}$$

Применение данного подхода позволило получить на открытой базе изображений лиц FG-NET [59] значение ошибки $MAE=5,16$ лет.

Лучшие результаты по мерам MAE и $CS(5)$ (см. параграф 1.6) получены при использовании метода *SVR* совместно с *накопительными* признаками (*Commulative Attributes*) [50]. Как показано в указанной работе, на общедоступной базе данных изображений лиц FG-NET [59] получены результаты $MAE=4,67$ и $CS(5) = 75\%$. Поскольку данный подход

представляется наиболее перспективным для дальнейшего развития, рассмотрим его более подробно.

Вновь рассмотрим обучающую выборку $G' \subset G$, заданную множеством пар прецедентов $\{\bar{x}_i, y_i\}$, $i = 1, n$; $G' \in \mathbb{R}^m$, \bar{x}_i - вектор признаков, соответствующий изображению i обучающей выборки G' , а y_i - соответствующее ему значение возраста; $Y \in \mathbb{R}$. Для каждого примера в данной обучающей выборке скалярное значение возраста преобразуется в вектор *накопительных* признаков \bar{a}_i . Вектор *накопительных* признаков \bar{a}_i имеет размерность r , соответствующую диапазону классифицируемых возрастов (y_{min}, y_{max}) . Если, например, рассматривается диапазон возрастов от 15 до 60 лет, то $r=60-15+1=46$. Каждое j -ое значение вектора *накопительных* признаков для i -го примера обучающей выборки $\{\bar{x}_i, y_i\}$ определяется следующим образом:

$$a_i^j = \begin{cases} 1, j \leq y_i - y_{min}; \\ 0, j > y_i - y_{min}. \end{cases}$$

Таким образом, в векторе *накопительных* признаков будет $(y_i - y_{min})$ значений 1, а все остальные значения будут равны 0. Сравним данные признаки с *ненакопительными* признаками \bar{c}_i , определёнными для i -го примера обучающей выборки $\{\bar{x}_i, y_i\}$ следующим образом:

$$c_i^j = \begin{cases} 1, j = y_i - y_{min}; \\ 0, j \neq y_i - y_{min}. \end{cases}$$

В векторе *ненакопительных* признаков есть только одно ненулевое значение: j -ый элемент принимает значение 1 в случае $j = y_i - y_{min}$. В этом и заключается ключевое отличие *накопительных* признаков от *ненакопительных*: в случае двух примеров с соседними значениям возрастов соответствующие им вектора *накопительных* признаков будут похожи, а в случае с различными значениями возрастов соответствующие им векторы *накопительных* признаков будут существенно различаться. Рассмотрим в качестве примера два элемента обучающей выборки со скалярными значениями возрастов 21 и 23. *Накопительные* и

ненакопительные признаки для данных элементов будут различаться в двух позициях. Теперь рассмотрим два элемента со значениями возрастов 21 и 42. *Ненакопительные* признаки для них будут различаться также в двух позициях, но *накопительные* - в 21. После получения *накопительных* признаков каждый элемент обучающей выборки G^l рассматривается как тройка $\{\bar{x}_i, \bar{a}_i, y_i\}$, где \bar{a}_i - соответствующий вектор *накопительных* признаков. Далее, в рассматриваемой работе ставится и решается задача определения зависимости между исходными признаками \bar{x}_i и *накопительными* \bar{a}_i , а так же между *накопительными* признаками и реальным значением возраста. Задача нахождения зависимости между исходными признаками и *накопительными* сводится к задаче гребневой регрессии [13].

Рассмотрим i -ый пример обучающей выборки G^l . Ему соответствуют \bar{x}_i - вектор исходных признаков, и \bar{a}_i - вектор *накопительных* признаков. Тогда, исходя из метода *SVM*, функционал потерь для j -го элемента *накопительных* признаков будет иметь вид [50]:

$$\frac{1}{2} \|\bar{u}^j\|^2 + C \sum_{i=1}^n \text{loss}(a_i^j, f^j(\bar{x}_i)) \rightarrow \min,$$

где $\text{loss}(\cdot)$ - обозначает функцию потерь, а $f^j(\bar{x}_i) = \langle \bar{u}^j, \bar{x}_i \rangle + u_0^j$.

Следовательно, совместное формирование признаков с помощью множественной регрессии формулируется следующим образом:

$$\sum_{j=1}^r \left[\frac{1}{2} \|\bar{u}^j\|^2 + C \sum_{i=1}^n \text{loss}(a_i^j, f^j(\bar{x}_i)) \right] \rightarrow \min.$$

В рассматриваемой работе [50] используется квадратичная функция потерь. Тогда функционал потерь принимает вид:

$$\frac{1}{2} \|\bar{U}\|^2 + C \sum_{i=1}^n \|\bar{a}_i^T - (\langle \bar{U}, \bar{x}_i \rangle + \bar{u}_0)\|_F^2 \rightarrow \min,$$

где $\bar{U} = [\bar{u}^1, \bar{u}^2 \dots \bar{u}^j \dots \bar{u}^r]$ – искомая матрица весовых коэффициентов размерностью $r \times n$, $\bar{a}_i = [a_i^1, a_i^2 \dots a_i^j \dots a_i^r]$ – обучающие векторы накопительных признаков, а r – количество значений возраста в интервале (y_{min}, y_{max}) . Для матриц используется норма Фробениуса [1], определяемая для матрицы D размером $m \times n$ как

$$\|D\|_F = \sqrt{\sum_{i=0}^n \sum_{j=0}^m d_{ij}^2}$$

Данная задача является задачей квадратичного программирования и имеет глобальное оптимальное решение в явно виде [50]:

$$\begin{pmatrix} \bar{U} \\ \bar{u}_0 \end{pmatrix} = -(Q^T Q)^{-1} Q^T P,$$

где Q и P – положительно полуопределённые матрицы, имеющие следующий вид:

$$P = \begin{pmatrix} -2C \sum_{i=1}^n \bar{x}_i \bar{a}_i^T \\ -2C \sum_{i=1}^n \bar{a}_i^T \end{pmatrix};$$

$$Q = \begin{pmatrix} 2C \sum_{i=1}^n \bar{x}_i \bar{x}_i^T + I & 2C \sum_{i=1}^n \bar{x}_i \\ 2C \sum_{i=1}^n \bar{x}_i & 2Cn \end{pmatrix}.$$

Управляющий параметр C подбирается экспериментально с помощью процедуры кросс-валидации [50].

Полученные значения матрицы весов \bar{U} позволяют перейти от исходных признаков \bar{x} к *накопительным* признакам \tilde{a}_i классифицируемых изображений:

$$\tilde{a}_i = (\langle \bar{U}, \bar{x}_i \rangle + \bar{u}_0). \quad (1.14)$$

Таким образом, осуществляется перевод исходных признаков в *накопительные*, которые учитывают взаимосвязь между соседними возрастами. Теперь обучающая выборка для i -го элемента представляется в виде $\{\tilde{a}_i, y_i\}$. Для нахождения зависимости между рассчитанными *накопительными* признаками и значением возраста используется метод *SVR*, в котором исходные признаки \bar{x}_i заменены на накопительные признаки \tilde{a}_i .

Отметим две важные идеи, связанные не с модернизацией самих методов классификации, а предусматривающие специальный порядок применения известных методов, что может способствовать повышению точности определения атрибута «возраст».

Первая идея изложена в работах [43, 65] и заключается том, что сначала производится предварительная классификация изображения лица человека по атрибуту «возраст» (в общем случае - любым доступным методом) и определяется некоторая возрастная группа, а затем, на втором шаге, строится регрессия внутри данной возрастной группы, по которой определяется значение возраста.

Вторая идея, изложенная в работе [66], предусматривает предварительную классификацию объектов по атрибуту «пол» (в общем случае - любым доступным методом), а затем решается задача классификации по атрибуту «возраст» внутри выделенной половой группы с использованием, указанной выше двухэтапной схемы определения значения возраста. Данные идеи представляются наиболее перспективными для дальнейшего развития и будут использованы в настоящей работе.

1.8. Итоги анализа методов определения атрибутов личности по изображению лица

Итоги анализа методов определения атрибутов личности («пол», «раса», «возраст») сведены в Табл. 1.2.

Анализ существующих методов определения атрибутов личности показал, что в последние годы одни из лучших результатов по рассмотренным в параграфе 1.6 критериям качества решающих функций получены на основе использования метода опорных векторов *SVM* для бинарной классификации, включая множественную классификацию на основе бинарной («один против всех»). При этом, для формирования вектора признаков используются как методы, основанные на выделении и анализе антропометрических точек (*ASM/AAM*), так и методы, основанные на анализе текстуры изображения (*LBP*). Для определения атрибута «возраст» лучшие результаты получены с использованием метода восстановления регрессии на основе опорных векторов (*SVR*) при формировании вектора признаков на основе антропометрических точек (*AAM*). Примеров использования для этой цели метода *LBP* в доступной литературе не обнаружено.

Таблица 1.2.

Основные методы определения атрибутов личности

Атрибут	Метод формирования вектора признаков изображения	
	<i>ASM/AAM</i>	<i>LBP</i>
«пол»	<i>SVM</i> [93, 62, 103]	<i>SVM</i> [84, 105] <i>Adaboost</i> [70, 105]
«раса»	<i>SVM</i> [62, 54]	<i>SVM</i> [74] <i>Adaboost</i> [111]
«возраст» (значение)	<i>SVR</i> [87, 65, 61]	-

Наиболее предпочтительными для использования и перспективными для дальнейшего развития представляются методы *SVM/SVR* на основе формирования вектора признаков изображения методом *LBP*. Целесообразность использования этого сочетания обосновывается следующими соображениями:

1) Метод *LBP* формирует вектор признаков изображения существенно большей размерности, чем метод *AAM*, что дает основание считать такое признаковое описание более информативным.

2) Метод *LBP* существенно проще в реализации, чем метод *AAM*, так как не требует сложного анализа изображения с целью выделения характерных точек на лице.

3) Метод опорных векторов *SVM* показывает наилучшие на сегодня результаты по точности классификации именно в сочетании с формированием вектора признаков изображения на основе метода *LBP*.

4) Регрессия на основе опорных векторов (*SVR*) показывает наилучшие результаты по определению возраста по метрикам *MAE* и *CS(ε)*.

Отсутствие в доступной литературе работ, использующих метод *SVR* в сочетании с *LBP* для определения атрибута «возраст», можно объяснить очень большой размерностью пространства признаков, порождаемой методом *LBP*, что затрудняет процесс восстановления регрессии. Однако высокая информативность признакового описания изображения, формируемого данным методом, позволяет предположить, что применение метода *LBP* для определения атрибутов личности позволит более точно решить данную задачу, а для снижения трудоемкости этой процедуры необходимо решить задачу понижения размерности пространства признаков путем исключения из вектора незначущих для классификации признаков.

Представляется весьма перспективной для дальнейшего развития идея сочетания нескольких подходов к определению атрибутов личности по изображению лица:

- сокращение размерности пространства признаков за счет удаления незначущих для классификации признаков;
- двухэтапная схема определения атрибута «возраст»;
- предварительное определение атрибута «пол» с последующей классификацией по атрибуту «возраст» внутри выделенной половой группы.

1.9. Выводы по главе 1

1) Анализ существующих методов признакового описания изображений показал, что наиболее информативным с точки зрения формирования вектора признаков, является метод локальных бинарных шаблонов (*LBP*). Этот метод целесообразно развивать в направлении его использования не только для определения атрибутов «пол» и «раса», но и для определения значения атрибута «возраст» с помощью восстановления регрессии.

2) При использовании метода локальных бинарных шаблонов (*LBP*) существенно возрастает размерность формируемого пространства признаков изображения. С одной стороны — это можно рассматривать как преимущество, так как сложные признаки содержат больше информации об изображении, но с другой стороны — большая размерность пространства признаков приводит к существенному усложнению задачи классификации. Поэтому, необходимы разработки в направлении снижения размерности пространства признаков, сформированных методом *LBP*, путем выделения наиболее информативных из них за счет использования априорной информации о человеческом лице. Данная

задача особенно актуальна при определении значения атрибута возраст с помощью восстановления регрессии.

3) Проведенный анализ существующих методов определения атрибутов личности по изображению лица показал, что на сегодняшний день не существует единого общепризнанного подхода к определению атрибутов, а в каждом случае используются (с разной степенью успешности) различные методы. Это усложняет процедуру определения атрибутов, если требуется определение не какого-то одного из них, а одновременно нескольких. Поэтому, целесообразно провести исследование в направлении разработки такого подхода к определению атрибутов, в рамках которого определение каждого необходимого атрибута является частью общего процесса, который может быть остановлен на нужном этапе. Наиболее перспективным для дальнейшего развития признан метод опорных векторов (*SVM*) для определения атрибутов «пол» и «раса», и регрессия на основе опорных векторов (*SVR*) для определения атрибута «возраст», использующие на этапе формирования вектора признаков метод локальных бинарных шаблонов (*LBP*).

2. РАЗРАБОТКА МЕТОДА ПОСЛЕДОВАТЕЛЬНОГО ОПРЕДЕЛЕНИЯ АТТРИБУТОВ ЛИЧНОСТИ ПО ИЗОБРАЖЕНИЮ ЛИЦА

2.1. Формализованное описание локальных бинарных шаблонов на базе принципа конечной топологии

2.1.1. Принцип конечной топологии

Рассмотрим основные положения принципа конечной топологии [32].

Пусть A – некоторое множество и 2^A – соответствующий ему булеан, т.е. множество всех его подмножеств. Некоторое конечное подмножество $T, T \subseteq 2^A$, будем называть конечной топологией на A , если выполнены следующие условия:

- 1) $\emptyset, A \in T$;
- 2) для любых A_1, A_2 имеет место $A_1 \cup A_2 \in T$;
- 3) для любых A_1, A_2 имеет место $A_1 \cap A_2 \in T$.

Множество A называется *носителем* топологии, а пара (A, T) – *топологическим пространством*.

Базой конечной топологии называется наименьшее ее подмножество B , объединением элементов которого можно представить любой непустой элемент $G, G \in T$ топологии.

Рассматривая конечную топологию как частично упорядоченное множество (с отношением теоретико-множественного включения), можно выделить ее *атомы* – элементы, доминирующие над \emptyset .

В настоящей работе рассматривается конечная топология, содержащая конечное количество атомов, объединения всевозможных подмножеств атомов, носитель топологии и пустое множество. Такую

топологию будем называть *атомарной* топологией. Атомами могут быть, например, элементы A_1, \dots, A_k некоторого разбиения A_1, \dots, A_n носителя A топологии.

Базу B атомарной топологии составляет множество B' атомов, дополненное носителем A топологии, если он не является объединением атомов, и пустым множеством \emptyset . Атомарная топология представляет собой дистрибутивную решетку, содержащую нижнюю булеву решетку, соответствующую булеану атомов и верхнюю двух- или одно-элементную решетку (включает объединение атомов и носитель топологии). Атомы топологии одновременно являются и атомами нижней решетки.

Именно нижняя булева решетка атомарной топологии и используется для представления свойств структурированного объекта (ниже она называется просто булевой решеткой конечной топологии). Атомы сопоставляются элементарным составляющим моделируемого объекта, и нижняя булева решетка используется в качестве области определения псевдо-булевой функции f , отражающей свойства элементарных составных частей и отношения, в которых они находятся в сочетании с другими элементами. Значениями этой функции на элементах решетки являются имена отношений, которым принадлежат соответствующие этим элементам подмножества атомов.

Совокупность объектов, имеющих одинаковые представления в виде функций рассмотренного вида, называется *топологической формой*.

Такое функциональное описание всех объектов изучаемой их совокупности позволяет путем сравнения функций, как это принято в теории построения тестов [27], выделять существенные для их различения подобласти решетки и определять метрику по числу несовпадений значений различных функций на них.

2.1.2. Топологические формы изображений

Определение топологической формы в предыдущем разделе можно расширить, допустив, что функции сравниваются не непосредственно по графикам, а по некоторым векторным характеристикам графиков. Такими характеристиками могут быть, например, гистограммы частот встречаемости в графиках различных значений функций. При этом появляется возможность сравнения топологических форм даже с различными областями определения.

Тогда топологической формой можно объявить совокупность объектов, имеющих близкие векторные описания.

Опишем этот подход при моделировании изображений.

Множество точек участка плоскости, на котором представлено изображение, объявим носителем A топологии (это непрерывное множество). Построим атомарную топологию на A . В качестве атомов выберем точки плоскости, соответствующие центрам пикселей. Базу топологии составляют атомы, пустое множество \emptyset и носитель топологии A . Нижняя решетка – это булеан множества атомов (включает также \emptyset), верхняя включает объединение атомов и носитель топологии.

Для создания топологической формы определим ряд псевдо булевых частичных функций на нижней решетке. Частичная функция $\varphi: X \rightarrow Y$ это множество функций, имеющих одинаковые значения на некотором подмножестве $X_\varphi \subseteq X$, задается графиком значений на множестве X_φ . Если X есть булева решетка, то частичная функция называется частичной псевдо булевой функцией. Псевдо булева функция $i(x,y)$ интенсивностей пикселей изображения определена на множестве атомов $\Pi = \{(x,y) \mid x \in \{1, \dots, N\}, y \in \{1, \dots, N\}\}$, и принимает на каждом из них целочисленные значения.

Определим 9-местное отношение на множестве атомов, соответствующее правилу формирования локальных бинарных шаблонов (см. Рис. 1.2):

$$\begin{aligned} \phi^9 = & \{((x_0, y_0), (x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), (x_5, y_6), (x_7, y_7), (x_8, y_8)) \setminus \\ & x_1 = x_7 = x_8 = x_0 - 1, \\ & x_2 = x_6 = x_0, \\ & x_3 = x_4 = x_5 = x_0 + 1, \\ & y_1 = y_2 = y_3 = y_0 + 1, \\ & y_4 = y_8 = y_0, \\ & y_5 = y_6 = y_7 = y_0 - 1 \}. \end{aligned}$$

Оператором *LBP* [103] на элементах отношения ϕ^9 определяется целочисленная функция f_{LBP} топологической формы. Она принимает на них значения, соответствующие бинарным 8-разрядным кодам (здесь “ \div ” – это операция «усеченного отрицания»: если $a > b$, то $b \div a = 0$, в остальных случаях $b \div a = b - a$):

$$\begin{aligned} f_{LBP}((x_0, y_0), (x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), (x_5, y_6), (x_7, y_7), (x_8, y_8)) = \\ = (\text{sign}(i(x_1, y_1) \div i(x_0, y_0)), \text{sign}(i(x_2, y_2) \div i(x_0, y_0)), \text{sign}(i(x_3, y_3) \div i(x_0, y_0)), \\ \text{sign}(i(x_4, y_4) \div i(x_0, y_0)), \text{sign}(i(x_5, y_5) \div i(x_0, y_0)), \text{sign}(i(x_6, y_6) \div i(x_0, y_0)), \\ \text{sign}(i(x_7, y_7) \div i(x_0, y_0)), \text{sign}(i(x_8, y_8) \div i(x_0, y_0))). \end{aligned}$$

Её так же можно рассматривать как частичную функцию на нижней решетке конечной топологии (она не определена на атомах, не имеющих соседа хотя бы по одной координате).

Сжатая функция топологической формы $\tilde{f}_{LBP}(x, y)$ определена на тех же атомах, присутствующих (в качестве значений первых координат) в элементах $((x_0, y_0), (x_1, y_1), \dots, (x_8, y_8))$ отношения ϕ^9 , которым соответствуют «равномерные» бинарные шаблоны (см. параграф 1.5), и принимает те же значения. Она не определена на прочих атомах и также может рассматриваться как частичная функция на нижней решетке конечной топологии. Гистограмма значений этой функции

рассматривается как статистическая модель изображения. Сравнением гистограмм можно группировать гистограммы, считая образующиеся группы топологическими формами.

Отметим, что выявление особой значимости *равномерных* бинарных шаблонов в работе [96] осуществлено чисто эмпирически. По принципу конечной топологии можно увидеть, что переход к *равномерным* шаблонам является примером изменения признакового пространства за счет использования информации об элементах внутренней структуры объекта (в данном случае – пикселей изображения, которые находятся в 9-местных отношениях 59 возможных типов (59-й тип включает неравномерные шаблоны)), Имена отношений 58 типов являются значениями частичной функции, определенной на множестве пикселей, участвующих в этих отношениях. Статистические характеристики такой функции используются в качестве векторной модели изображения. Подобный подход был успешно использован в работах [4, 5, 17, 18] для распознавания и классификации текстовой информации.

Предложенное на основе принципа конечной топологии формализованное описание локальных бинарных шаблонов и *равномерных* бинарных шаблонов отличает данное описание от принятого чисто эмпирического их определения и является предпосылкой для унифицированного функционального представления изображений.

2.2. Снижение размерности пространства признаков изображения для определения атрибутов личности

Анализ методов формирования вектора признаков, описывающего нормализованное изображение лица человека, и алгоритмов классификации изображений лиц с целью определения таких атрибутов личности, как «пол» и «раса», показал, что одни из лучших на сегодня

результатов были получены при использовании метода локальных бинарных шаблонов (*LBP*) в сочетании с методом опорных векторов (*SVM*). Однако распространение данного подхода на задачу определения атрибута «возраст» с помощью регрессии на основе опорных векторов (*SVR*) затруднено из-за очень высокой размерности пространства признаков, которое формирует метод *LBP* (около 2500). Рассмотрим возможности снижения размерности данного пространства за счет использования априорной информации об изображении лица человека. В качестве исходных предпосылок для успешного решения этой задачи выделим следующие:

- 1) симметричность лица;
- 2) различная информативность отдельных участков изображения лица.

Первая предпосылка является весьма очевидной, однако использование этого априорного факта в литературе не обнаружено. Анализируя алгоритм формирования пространства признаков методом *LBP* (см. параграф 1.5) легко видеть, что итоговая гистограмма будет содержать практически идентичные пары гистограмм, построенные по симметричным участкам лица (см. Рис. 2.1). Поэтому, удаление половины изображения (половины гистограмм из признакового описания) никак не повлияет на информативность вектора признаков (трудно предположить, что пол, раса или возраст левой половины изображения лица будут отличаться от значения этих атрибутов для правой половины).

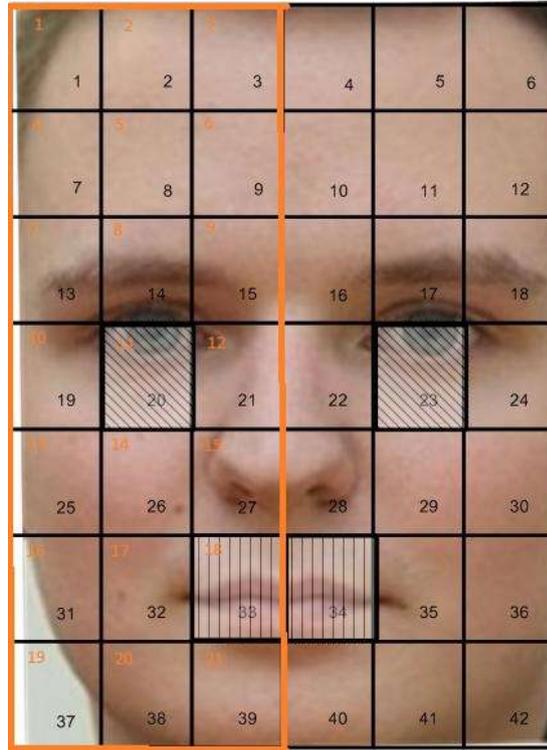


Рис. 2.1. Изображение с примерами практически идентичных *LBP*-регионов

Использование данного очевидного факта приводит к снижению размерности пространства признаков в 2 раза без снижения его информативности. Тогда при рассмотрении только *равномерных* бинарных шаблонов (см. параграф. 1.5.) и разбиении исходного нормализованного изображения на 42 региона размерность пространства снижается с 2436 до 1218, так как число регионов равно 21. В этом случае вектор признаков изображения \bar{x}_i ($i=1,n$) принимает вид (с учетом использования *равномерных* бинарных шаблонов, см. параграф 1.5):

$$\Psi_{LBP_{\text{равн.}}}(i) = \bar{x}_i = (\bar{0}_{i_1}, \bar{1}_{i_1}, \bar{2}_{i_1}, \dots, \bar{x}_{i_1}, \dots, \bar{57}_{i_1}, \dots, \bar{0}_{i_j}, \dots, \bar{x}_{i_j}, \dots, \bar{57}_{i_j}, \dots, \bar{0}_{i_{21}}, \dots, \bar{x}_{i_{42}}, \dots, \bar{57}_{i_{21}}), \quad (2.1)$$

где \bar{x} по-прежнему десятичный числовой эквивалент *равномерного* шаблона с порядковым номером x , а \bar{x}_{i_j} – частота встречаемости такого шаблона среди вхождений *равномерных* шаблонов в полном векторном представлении j -го фрагмента половины i -го изображения при упорядоченности *равномерных* шаблонов по их числовым эквивалентам.

Следует отметить следующее дополнительное преимущество использования априорной информации о симметричности лица. Как отмечалось ранее (см. параграф 1.3), на этапе нормализации изображение лица подвергается аффинному преобразованию с целью получения фронтального изображения из реального, которое практически всегда несколько повернуто. При этом наиболее трудности с качеством изображения возникают при восстановлении дальней (частично скрытой) половины лица (см. Рис. 2.2). Предлагаемая идея использования только одной половины изображения делает вообще не нужной процедуру восстановления второй половины, что сокращает затраты времени на процесс нормализации изображения без снижения качества результата.



Рис. 2.2. Формирование фронтального изображения лица из реального (аффинное преобразование) с выделением половины изображения

Вторая предпосылка к снижению размерности пространства признаков (различная информативность участков изображения лица) так же представляется вполне очевидной. Да и внутри региона различные участки изображения содержат разное количество информации, скрытой в бинарных шаблонах (частично данный факт отражен в *равномерных* бинарных шаблонах). Поэтому из итогового вектора признаков

желательно удалить те признаки, которые являются мешающим классификации шумом.

Вновь рассмотрим обучающую выборку по атрибуту «пол» $G^l \subset G$, заданную множеством пар прецедентов $\{\bar{x}_i, y_i\}$, $i = 1, n$; $G^l \in \mathbb{R}^m$, \bar{x}_i - вектор признаков изображения i обучающей выборки G^l , полученный по половине исходного изображения (см. соотношение 2.1), а y_i - соответствующее ему значение пола, $y_i \in \{-1, +1\}$.

Для простоты обозначения пронумеруем все признаки от 1 до 1218, т.е.

$$\Psi_{LBP_{\text{равн.}}} (i) = \bar{x}_i = (\bar{0}_{i_1}, \bar{1}_{i_1}, \bar{2}_{i_1}, \dots, \bar{x}_{i_1}, \dots, \bar{57}_{i_1}, \dots, \bar{0}_{i_j}, \dots, \bar{x}_{i_j}, \dots, \bar{57}_{i_j}, \dots, \bar{0}_{i_{21}}, \dots, \bar{x}_{i_{21}}, \dots, \bar{57}_{i_{21}}) = (x_i^1, x_i^2, x_i^3 \dots x_i^{1218}).$$

Теперь x_i^j – это частота встречаемости j -го по порядку признака ($j=1, 1218$) десятичного эквивалента соответствующего *равномерного* бинарного шаблона среди вхождений таких шаблонов в полном векторном представлении для половины i -го изображения.

Для автоматического решения данной задачи – выделение наиболее информативных признаков из исходных – предлагается использовать алгоритм *Adaboost* для бинарной классификации по атрибуту «пол», рассмотренный в параграфе 1.7.1.

Основная идея заключается в том, чтобы применить процедуру бустинга к каждому из признаков, т.е. заставить алгоритм *Adaboost* принимать решение об отнесении изображения к тому или иному классу по каждому единичному признаку x_i^j отдельно, $i = 1, n$; $j=1, 1218$. В этом случае задача классификации с помощью алгоритма *Adaboost* может быть сформулирована следующим образом (см. параграф 1.7.1):

$$f(x^j) = \text{sign}\left(\sum_{i=1}^n \alpha^j b(x_i^j)\right), \quad (2.2)$$

т.е. решающая функция будет сформирована таким образом, чтобы классифицировать изображения только по j -тому признаку ($j=1, 1218$) всей обучающей выборки G^l . В данном соотношении весовые коэффициенты α^j определяют значимость базовой функции $b(x)$, в качестве которой предлагается использовать простейшие бинарные классификаторы с разделением по пороговому значению (*Decision Stump*). Данное пороговое значение вычисляется, как среднее значение параметра x_i^j по всей обучающей выборке G^l .

Функционал потерь $L(f(x^j), G^l)$ определим как число ошибок, допускаемых комитетом из 1218 решающих базовых функций (по числу признаков) на признаках x^j всех объектов обучающей выборки G^l :

$$L(f(x^j, G^l)) = \sum_{j=1}^n \frac{1 - z_j}{2};$$

$$z_j = y_j \sum_{i=1}^n \alpha^j b_i(x_i^j).$$

В соответствии с алгоритмом *Adaboost* вместо пороговой функции $z < 0$ будем использовать ее непрерывно дифференцируемую оценку сверху вида $E(z) = \exp(-z)$:

$$L(f(x^j, G^l)) \leq \tilde{L}(f(x^j, G^l)) = \frac{1}{2} \sum_{i=1}^n [1 - \exp(-\alpha^j b(x_i^j) y_i)].$$

Минимизация данного функционала приводит к следующему результату для весовых коэффициентов (см. параграф 1.7.1):

$$\alpha^j = \frac{1}{2} \ln \frac{P(x^j)}{N(x^j)}, \quad (2.3)$$

где $P(x^j)$ - доля правильных классификаций, а $N(x^j)$ - доля ошибочных классификаций по признаку x^j на всех объектах обучающей выборки: $P(x^j) > N(x^j)$, $P(x^j) + N(x^j) = 1$. Для признаков x^j , у которых $P(x^j) < N(x^j)$, коэффициенты $\alpha^j = 0$. Полученные весовые коэффициенты нормируются так, чтобы их сумма была равна 1.

Результатом применения данного подхода будет совокупность весовых коэффициентов α^j , определяющих значимость базового алгоритма $b(x)$. Поскольку базовый алгоритм применялся к единственному признаку x^j , данный весовой коэффициент будет характеризовать значимость конкретного признака x^j при принятии решения об отнесении объекта к определенному классу.

Далее предлагается упорядочить признаки по убыванию весовых коэффициентов, и ввести порог значимости δ . Для решения задачи классификации предлагается использовать только те признаки, для которых выполняется условие:

$$\sum_{j=1}^N \alpha^j < \delta,$$

т.е. оставить только первые N признаков, суммарный вес которых составляет, например, 95%. Данные признаки будем называть *усиленными* и обозначать \hat{X} , а формализованную обучающую выборку, содержащую только *усиленные* признаки - \hat{G}^l . Отметим, что при выполнении предлагаемой процедуры необходимо запомнить значимые *равномерные* бинарные шаблоны для каждого региона изображения, чтобы выделять только эти шаблоны из изображений в процессе формирования вектора признаков (см. параграф 1.5).

Таким образом, в каждом регионе изображения лица будут оставлены только те бинарные шаблоны, суммарная значимость которых выше заданного порога. Как будет показано в главе 3, при исследовании

реальных изображений данный подход, даже при значении $\delta = 0,95$ (оставляем признаки, имеющие хоть какую-нибудь значимость для классификации), позволяет сократить размерность пространства признаков примерно до 280, что уже соизмеримо с размерностью пространства признаков, формируемого при использовании существенно более трудоемкого метода *ААМ*.

Следовательно, использование априорной информации о человеческом лице в совокупности с автоматическим определением значимости признаков изображения на основе процедуры бустинга, позволяет почти в 10 раз снизить размерность вектора признаков изображения, полученного методом *LBP*, и довести размерность пространства признаков примерно до 280 (при значении порога значимости $\delta = 0,95$).

Отметим еще одну особенность формирования вектора признаков изображения методом *LBP*. Как было показано в параграфе 1.7.1, для обеспечения линейной разделимости классов используют идею *спрямляющих* пространств, т.е. от исходного пространства признаков \mathbf{X} переходят к пространству признаков \mathbf{H} большей размерности за счет применения некоторого функционального преобразования, что было актуально при использовании метода *ААМ*, формирующего относительно небольшое количество признаков. В рассматриваемом случае размерность пространства признаков $\hat{\mathbf{X}}$ даже после его сокращения предложенным выше способом остается весьма высокой, что позволяет отказаться от применения данной процедуры и выполнять все операции по классификации в исходном пространстве. Это также сокращает трудоемкость процедур обучения и классификации. Конкретные результаты по затратам времени на бинарную классификацию будут приведены в главе 3.

Выделим важный результат, полученный в данном параграфе: предложенный подход может быть применен к другим способам формирования пространства признаков. Если в будущем для формирования признаков будет предложен более удачный, чем *LBP*, метод, то к нему может быть применена предложенная процедура, что позволил выделить из всего пространства признаков, только те, которые значимы для решения задачи классификации.

Полученные признаки названы *усиленными*, и именно их предлагается использовать для определения других искомым атрибутов личности – расы и возраста.

2.3. Модификация метода опорных векторов для бинарной классификации изображений при определении атрибутов «пол» и «раса»

Как было показано ранее (см. параграфы 1.7.1 и 1.7.2), классификация изображений по атрибутам «пол» и «раса» является бинарной или может быть сведена к бинарной. В настоящем параграфе предлагается модификация метода опорных векторов (*SVM*), с помощью которого получены одни из лучших на сегодня по критерию *Accuracy* (соотношение (1.3)) результаты при определении атрибутов «пол» и «раса».

Прежде всего, отметим, что сокращение пространства признаков изображения, предложенное в предыдущем параграфе, позволяет существенно сократить трудоемкость процедуры классификации методом *SVM* при формировании вектора признаков изображения методом *LBP* с повышением качества классификации (конкретные результаты будут приведены в главе 3).

В качестве исходного алгоритма для модификации процедуры обучения был взят метод *SVM* с «мягким» зазором (см. параграф 1.7.1), в котором алгоритму разрешено допускать ошибки (случай линейно неразделимой выборки), но величина этих ошибок минимизируется. Предлагаемая модификация основана на применении процедуры бутстраппинга (*Bootstrapping*) - обучение на «трудных» примерах [69, 79].

Необходимо отметить, что данная процедура традиционно используется при решении другой задачи из области компьютерного зрения - идентификации заданного объекта на изображении, содержащем множество самых разнообразных (похожих) объектов [69] (например, выделение пешехода на изображении улицы, на котором может находиться множество людей в автомобилях, на велосипедах и на рекламных плакатах). Основной проблемой при решении этой задачи является огромное количество примеров фона (не пешеходов) по сравнению с искомыми объектами. Поэтому центральной идеей бутстраппинга является включение в обучающую выборку так называемых «трудных» примеров, т.е. таких примеров фона, которые ошибочно идентифицируются классификатором как искомые объекты.

Применительно к рассматриваемой задаче определения атрибутов личности по изображению лица идея использования процедуры бутстраппинга состоит в следующем. Пусть, по-прежнему, имеется обучающая выборка $\hat{G}^l \subset G$, заданная множеством пар прецедентов $\{\hat{x}_i, y_i\}$, $i = 1, n$; $\hat{x}_i \in \mathbb{R}^m$, $y_i \in \{-1, +1\}$, где вектор признаков изображения i содержит только *усиленные* признаки, формирование которых рассмотрено в предыдущем параграфе.

Предлагаемая процедура бутстраппинга содержит три шага:

1) Обучающая выборка \hat{G}^l случайным образом делится на две неравные части \hat{G}_1^l и \hat{G}_2^l в соотношении 1 к 2: $\hat{G}_1^l \cap \hat{G}_2^l = \emptyset$. Первая часть \hat{G}_1^l используется в качестве обучающей выборки для формирования

методом *SVM* с «мягким» зазором предварительной оценки \tilde{f} решающей функции f .

2) Вторая часть \hat{G}_2^l используется в качестве тестовой выборки (обучающее тестирование), на которой оценивается предсказательная способность решающей функции \tilde{f} . При этом выделяются все объекты $\hat{x}_i \in \hat{G}_2^l$, которые были неправильно классифицированы.

3) Ошибочно классифицированные объекты, выделенные на этапе 2 («трудные» для классификации примеры), добавляются к обучающей выборке \hat{G}_1^l , т.е. формируется расширенная «трудными» примерами новая обучающая выборка \hat{G}_{Inew}^l . По этой расширенной выборке методом *SVM* с «мягким» зазором производится переобучение, т.е. окончательно формируется решающая функция f .

Важным условием применимости предложенной процедуры является большая мощность исходной обучающей выборки (тысячи объектов). К счастью, для рассматриваемой задачи это условие является легко выполнимым, так как формирование необходимого количества изображений лиц людей с известными атрибутами с помощью открытых источников в сети Интернет не является проблемой (этот вопрос будет рассмотрен в главе 3).

Отметим следующую особенность метода *SVM* с «мягким» зазором [11]. Как было отмечено в параграфе 1.7.1, методу разрешено ошибаться в процессе формирования разделяющей классы гиперплоскости (см. Рис. 1.10), минимизируя отклонение (ошибку) ξ от разделяющей гиперплоскости. При этом использование процедуры бутстраппинга (добавления «трудных» примеров в обучающую выборку) может привести к тому, что решающая функция f (разделяющая классы гиперплоскость) будет излишне настраиваться на эти «трудные» примеры, которые могут носить случайный характер, т.е. быть «шумом». Такое ложное обучение может ухудшить предсказательную способность решающей функции f , так

как «зашумленные» обучающие объекты могут не содержать значимой для классификации информации.

Для снижения вероятности ложного обучения предлагается ввести следующую модификацию в процедуру бутстраппинга применительно к методу *SVM* с «мягким» зазором.

Пусть при использовании процедуры бутстраппинга на втором этапе рассмотренного выше алгоритма были выделены «трудные» примеры $\hat{x}_i \in \hat{G}_2^l$, на которых решающая функция допустила ошибку. Выделим среди этих ошибочно классифицированных объектов два объекта разных классов $\hat{x}_{max(+1)}$ и $\hat{x}_{max(-1)}$, имеющие максимальное расстояние $\xi_{max(+1)}$ и $\xi_{max(-1)}$ от соответствующей границы зазора между классами (см. Рис. 2.3), и исключим эти объекты («самые трудные» для классификации примеры) из множества «трудных» примеров, которые будут добавлены к обучающей выборке \hat{G}_1^l на третьем этапе процедуры бутстраппинга при формировании новой обучающей выборки \hat{G}_{new}^l .

Предложенную процедуру будем называть алгоритмом *усеченного* бутстраппинга. В качестве обоснования целесообразности применения данного алгоритма приведем следующее рассуждение. Если «самые трудные» для классификации примеры $\hat{x}_{max(+1)}$ и $\hat{x}_{max(-1)}$ действительно являются «шумом» и не должны учитываться методом *SVM* с «мягким» зазором в процессе формирования решающей функции f , то их исключение из расширенной обучающей выборки \hat{G}_{new}^l является обоснованным.

Если же данные объекты являются типичными представителями соответствующих классов, то в обучающей выборке \hat{G}_{new}^l непременно будут присутствовать аналогичные объекты (выборки \hat{G}_1^l , \hat{G}_2^l были сформированы случайным образом и содержат тысячи объектов), и исключение этих двух практически не повлияет на итоговый результат.

Реализация предложенной модификации приводит к следующим расчетным соотношениям, основанным на использовании *усиленных* признаков изображений, формирование которых было рассмотрено выше (см. параграф 2.1). При этом рассматривается первая часть обучающей выборки \hat{G}_1^l , содержащая n объектов.

Задача оптимизации при линейной норме вектора «мягкого» отступа ξ_i формулируется следующим образом (см. параграф 1.7.1):

$$\begin{cases} \left\{ \frac{1}{2} \|\bar{w}\|^2 + C \sum_{i=1}^n \xi_i \right\} \rightarrow \min_{w, w_0}; \\ y_i (\langle \bar{w}, \hat{x}_i \rangle + w_0) \geq 1 - \xi_i; \\ \xi_i \geq 0, \end{cases} \quad (2.4)$$

где ξ_i - величина ошибки.

Лагранжиан данной задачи имеет вид:

$$\mathcal{L} = \frac{1}{2} \|\bar{w}\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \lambda_i (y_i (\langle \bar{w}, \hat{x}_i \rangle + w_0) - 1 + \xi_i) - \sum_{i=1}^n \eta_i \xi_i.$$

Приравняв к нулю производные, получим систему уравнений:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \bar{w}} &= \bar{w} - \sum_{i=1}^n \lambda_i y_i \hat{x}_i = 0; \\ \frac{\partial \mathcal{L}}{\partial \xi_i} &= \bar{w} - C - \lambda_i - \eta_i = 0; \\ \frac{\partial \mathcal{L}}{\partial w_0} &= \sum_{i=1}^n \lambda_i y_i = 0. \end{aligned}$$

Подставив решения этих уравнений в прямую задачу, получаем двойственное представление задачи в виде максимизации функционала:

$$\mathcal{L} = \sum_{i=1}^n \lambda_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j y_i y_j \langle \hat{x}_i, \hat{x}_j \rangle \rightarrow \max \quad (2.5)$$

при условии

$$\sum_{i=1}^n \lambda_i y_i = 0;$$

$$0 \leq \lambda_i \leq C.$$

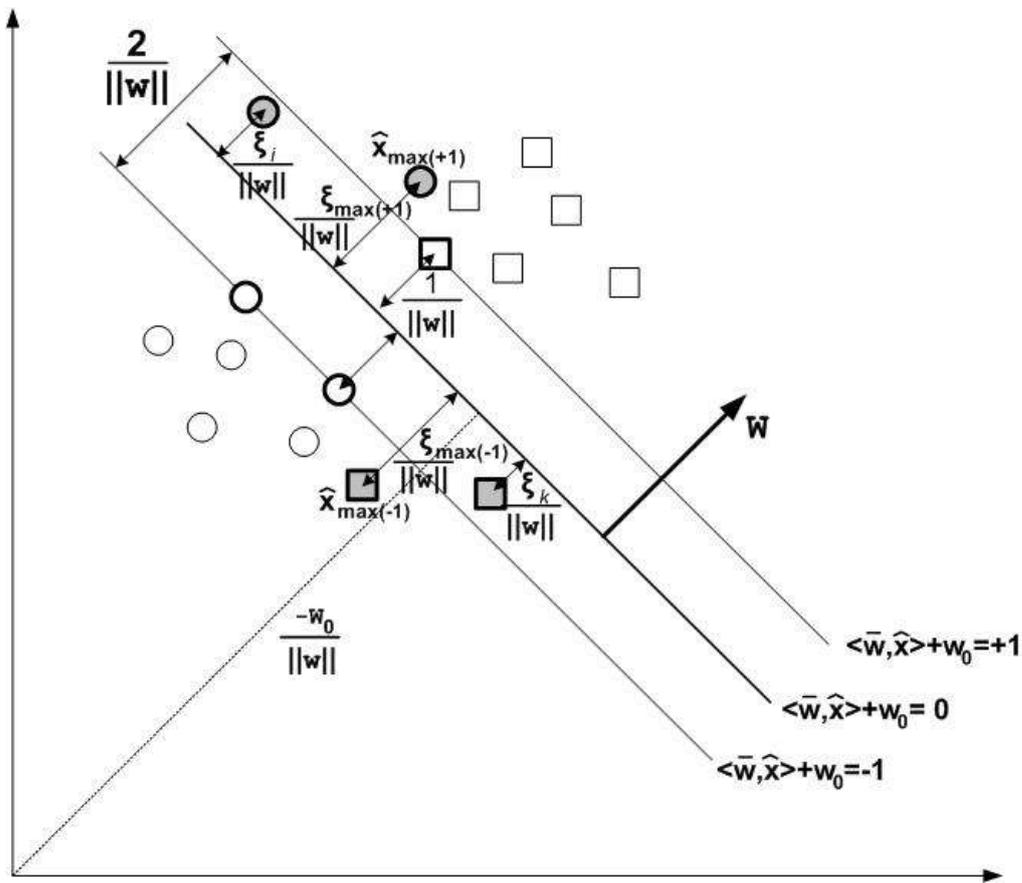


Рис.2.3. Выделение «самых трудных» примеров в процедуре усеченного бутстраппинга

Условия Каруша-Куна-Такера [11] имеют вид:

$$\lambda_i (y_i (\langle \bar{w}, \hat{x}_i \rangle + w_0) - 1 + \xi_i) = 0;$$

$$(\lambda_i - C) \xi_i = 0; i = 1, n.$$

Из условий Каруша-Куна-Такета и граничных условий оптимизационной задачи следует:

- если $y_i(\langle \bar{w}, \hat{x}_i \rangle + w_0) > 1$, то $\lambda_i = 0$ и $\xi_i = 0$ (правильная классификация);
- если $y_i(\langle \bar{w}, \hat{x}_i \rangle + w_0) < 1$, то $\lambda_i = C$ и $\xi_i > 0$ (ошибочная классификация);
- если $y_i(\langle \bar{w}, \hat{x}_i \rangle + w_0) = 1$, то $0 \leq \lambda_i \leq C$ и $\xi_i = 0$ (вектор является опорным).

Решающая функция для классификации произвольного вектора признаков \hat{x} принимает вид:

$$f(\hat{x}) = \text{sign}\left(\sum_{i=1}^n \lambda_i y_i \langle \hat{x}_i, \hat{x} \rangle + w_0\right), \quad (2.6)$$

а w_0 находится как медиана по всем опорным векторам:

$$w_0 = \text{med}(\langle \bar{w}, \hat{x}_i \rangle - y_i).$$

Положительная константа C является управляющим параметром метода и позволяет находить компромисс между максимизацией ширины разделяющей классы полосы и минимизацией суммарной ошибки. Данная константа C подбирается итерационно так, чтобы разделяющая классы гиперплоскость разделяла элементы обучающей выборки с минимальным значением нормы вектора [11, 13].

С помощью полученной решающей функции производится обучающее тестирование по обучающей выборке \hat{G}^l и выделяются все ошибки классификации. Среди этих ошибочно классифицированных векторов («трудных» примеров) выделяются два «самых трудных», для которых значения ξ_i принимают максимальные значения.

Далее, в соответствии с предложенным алгоритмом *усеченного* бутстраппинга формируется расширенная обучающая выборка \hat{X}_{new}^l для которой повторно производятся приведенные расчеты с целью определения уточненной решающей функции.

Как было указано ранее (см. параграф 1.7.2) для определения атрибута «раса» по изображению лица используется множественная классификация по схеме «один против всех», в соответствии с которой сначала в результате предложенной модификации бинарной классификации отделяются объекты, относящие к классу «европеиды», а потом оставшиеся объекты разделяются модифицированной бинарной классификацией на «монголоидов» и «негроидов».

Необходимо отметить, что предложенная процедура *усеченного* бутстраппинга может быть применена к другим методам формирования решающей функции, т.е. может быть применена к любому другому методу, который появится в будущем, для его модификации с целью повышения качества классификации.

2.4. Модификация двухэтапной схемы определения атрибута «возраст» методом восстановления регрессии на основе опорных векторов

На основе проведенного анализа различных подходов к определению атрибута «возраст» по изображению лица путем решения задачи восстановления регрессии предлагается следующая последовательность действий для решения данной задачи, представляющая собой сочетание самых лучших на сегодня идей (см. параграф 1.7.3):

- использование метода восстановления регрессии на основе опорных векторов (*SVR*) [61, 65, 87];

- использование идеи *накопительных* признаков [50] в качестве исходных данных для метода *SVR*;
- использование двухэтапной схемы определения атрибута «возраст» [61, 65]; на первом этапе осуществляется вычисление приближенного значения возраста по решающей функции $f_0(\hat{x})$, определенной на основе регрессии, построенной по обучающей выборке \hat{G}^l по всему диапазону возрастов (y_{min}, y_{max}) ; на втором этапе производится уточнение данного значения за счет использования решающей функции $f_d(\hat{x})$, сформированной на основе регрессии, построенной по конкретной возрастной группе d , в которую попало значение возраста, определенное на первом этапе; как и в работе [65], будем называть используемую на первом этапе регрессию глобальной, а на втором этапе - локальной.

Как будет показано в главе 3, такое объединение известных подходов в единый алгоритм повышает точность определения атрибута «возраст» по сравнению с независимым применением каждого из них. В настоящей работе предлагается усовершенствовать двухэтапную схему определения возраста.

В соответствии с базовым алгоритмом [61], весь интервал возрастов (y_{min}, y_{max}) делится на непересекающиеся диапазоны, по которым на основе локальной регрессии определяются решающие функции для каждого из них. Поскольку данные диапазоны были фиксированными, это вызывало повышенную погрешность в определении возраста при попадании приближенного значения возраста в область вблизи границы диапазона. В работе [77] данную проблему предложено решать путем использования пересекающихся диапазонов. При этом решающая функция строится на основе локальной регрессии, формируемой по пересекающимся диапазонам возрастов.

В настоящей работе предлагается отказаться от фиксированных значений границ диапазонов и использовать «плавающие» границы, зависящие от конкретного значения возраста, полученного на первом этапе.

Главной предпосылкой для предлагаемой модификации является следующее утверждение: значимость ошибки в определении возраста зависит от абсолютного значения возраста. Так, например, ошибка в 2 года при определении возраста 15-ти летнего подростка представляется весьма значимой, в то время как ошибку в 5 лет при определении возраста 70-тилетнего человека вряд ли можно считать существенной.

При этом на втором этапе для проведения процедуры локальной регрессии предлагается использовать диапазон возрастов $(y - d^-(y), y + d^+(y))$, где y – приближенное значение возраста, определенное на первом этапе. Заметим, что в общем случае, $d^-(y) \neq d^+(y)$, т.е. диапазон может быть не симметричным. Такой выбор диапазона целесообразен в том случае, когда количество прецедентов существенно неравномерно распределено по возрастам.

Вторым направлением для модификации двухэтапной схемы определения возраста является изменение функции потерь, используемой при решении задачи восстановления регрессии (см. параграф 1.7.3). В известных на сегодня методах [65, 66] используется ε -чувствительная функция потерь с фиксированным значением чувствительности ε . По аналогии с предложенной выше идеей «плавающих» границ диапазонов, на втором этапе двухэтапной схемы определения возраста предлагается использовать разное значение чувствительности функции потерь в зависимости от абсолютного значения возраста, т.е. значение $\varepsilon = \varepsilon(y)$ зависит от значения возраста y , полученного на первом этапе. Графическая иллюстрация идеи предложенных модификаций показана на Рис. 2.4.

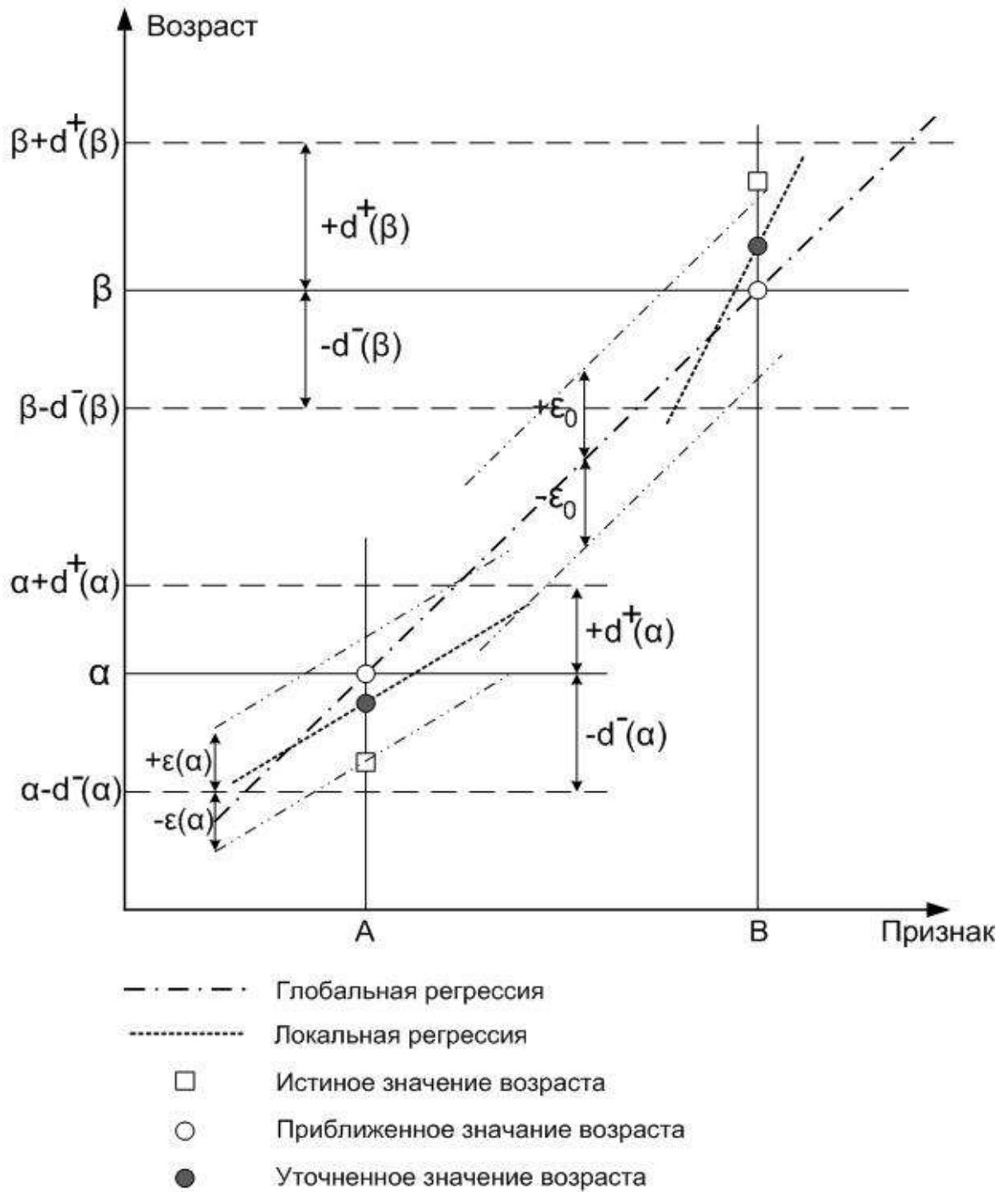


Рис. 2.4. Иллюстрация идеи использования «плавающих» границ диапазонов и зависимости чувствительности ε функции потерь от приближенного значения атрибута «возраст»

Наконец, в качестве третьей модификации для определения атрибута «возраст» предлагается использовать процедуру «усеченного» бутстраппинга, разработанную в предыдущем параграфе, при формировании как решающей функции на основе глобальной регрессии, так и при формировании решающих функций на основе локальных регрессий внутри каждого диапазона. Это означает, что при расширении обучающей выборки «трудными» примерами, полученными на этапе обучающего тестирования, два объекта, значение возраста для которых было определено с максимальной ошибкой, не используются при переобучении решающей функции.

В итоге, предлагается следующая процедура формирования совокупности решающих функций для задачи определения возраста на основе метода *SVR*, основанная на полученных ранее результатах и предложенных модификациях двухэтапной схемы определения атрибута «возраст».

- 1) Формируется множество усиленных признаков \hat{X}^l для формализованной обучающей выборки \hat{G}^l в соответствии с предложенным выше алгоритмом (см. параграф 2.1);
- 2) В соответствии с методом формирования накопительных признаков [50] (см. параграф 1.7.3) по усиленным признакам \hat{X}^l определяются накопительные признаки \hat{A}^l ;
- 3) По полученным *накопительным* признакам \hat{A}^l формализованной обучающей выборки \hat{G}^l решается задача восстановления регрессии методом *SVR* по всей обучающей выборке с заданным значением ε -чувствительности ε_0 , т.е. формируется решающая функция $f_0(\hat{a}, \varepsilon_0, \hat{G}^l)$, для предварительного определения значение возраста \tilde{y}_i (этап глобальной регрессии двухэтапной схемы определения возраста), т.е. $\tilde{y}_i = f_0(\hat{a}, \varepsilon_0, \hat{G}^l)$, $i=1, n$. При этом, используется

схема «усеченного» бутстраппинга (см. параграф 2.2) для всей обучающей выборки:

- выборка \hat{G}^l случайным образом делится на две части \hat{G}_1^l и \hat{G}_2^l в соотношении 1 к 2;
- по выборке \hat{G}_1^l методом **SVR** формируется решающая функция;
- по выборке \hat{G}_2^l проводится обучающее тестирование, и выделяются «трудные» примеры («трудными» считаются примеры, не попавшие в полосу чувствительности ε_0);
- из «трудных» примеров удаляются два «самых трудных» примера, имеющих максимальную ошибку, и формируется выборка \hat{G}_{Inew}^l ;
- по выборке \hat{G}_{Inew}^l окончательно определяется решающая функция $f_0(\hat{a}, \varepsilon_0, \hat{G}^l)$.

4) Для каждого значения возраста y выделяется диапазон возрастов $(y - d^-(y), y + d^+(y))$, с значениями отступа, зависящими от y , которые определяются экспертом исходя из специфики решаемой задачи; наиболее очевидная идея – чем больше значение возраста, тем шире диапазон, но возможны и другие правила.

5) Внутри каждого диапазона по соответствующему подмножеству обучающей выборки $\hat{G}_d^l: y_i \in (y - d^-(y), y + d^+(y))$, решается задача восстановления регрессии методом **SVR**, в которой чувствительность функции потерь $\varepsilon(y)$ зависит от значения y и так же определяется экспертом, исходя из специфики решаемой задачи. По-прежнему, наиболее очевидная идея – чем больше значение y , тем ниже чувствительность, т.е. тем больше $\varepsilon(y)$. Таким образом, формируется решающая функция $f_y(\hat{a}, \varepsilon(y), d(y), \hat{G}^l)$ для определения уточненного значения возраста. При этом, как и на шаге 4, используется схема «усеченного» бутстраппинга (см. параграф 2.3) для каждого диапазона

$(y - d^-(y), y + d^+(y))$, обучающей выборки («трудными» по-прежнему считаются примеры, не попавшие в полосу чувствительности $\varepsilon(y)$).

Итогом реализации предложенного алгоритма является совокупность решающих функций, включающая в себя функцию $f_0(\hat{a}, \varepsilon_0, \hat{G}^l)$, определяющую приближенное значение возраста y по всему диапазону возрастов (глобальная регрессия), и множество функций $f_y(\hat{a}, \varepsilon(y), d(y), \hat{G}^l)$, определяющих уточненное значение возраста на втором этапе (локальная регрессия) двухэтапной схемы; $y \in (y_{min}, y_{max})$.

Реализация предложенной процедуры приводит к следующим расчетным соотношениям.

Прежде всего, решается задача перехода от *усиленных* признаков $\hat{\mathbf{X}}^l$ к *накопительным* признакам $\hat{\mathbf{A}}^l$. Как было показано ранее (см. параграф 1.7.3), решение данной задачи при квадратичной функции потерь сводится к минимизации следующего функционала:

$$\frac{1}{2} \|\bar{u}\|^2 + C \sum_{i=1}^n \|\hat{a}_i^T - (\langle \bar{u}, \hat{x}_i \rangle + \bar{u}_0)\|_F^2 \rightarrow \min,$$

где $\bar{u} = [\bar{u}^1, \bar{u}^2 \dots \bar{u}^j \dots \bar{u}^r]$ – искомая матрица весовых коэффициентов размерностью $r \times n$, $\hat{a}_i = [a_i^1, a_i^2 \dots a_i^j \dots a_i^r]$ – обучающие векторы *накопительных* признаков, а r – количество значений возраста в интервале (y_{min}, y_{max}) . Для матриц используется норма Фробениуса, определяемая для матрицы \mathbf{D} размером $m \times n$ как

$$\|\mathbf{D}\|_F = \sqrt{\sum_{i=0}^n \sum_{j=0}^m d_{ij}^2}$$

Данная задача является задачей квадратичного программирования и имеет глобальное оптимальное решение в явном виде [50]:

$$\begin{pmatrix} \bar{u} \\ \bar{u}_0 \end{pmatrix} = -(Q^T Q)^{-1} Q^T P,$$

где Q и P - положительно полуопределённые матрицы, имеющие следующий вид:

$$P = \begin{pmatrix} -2C \sum_{i=1}^n \hat{x}_i \bar{a}_i^T \\ -2C \sum_{i=1}^n \bar{a}_i^T \end{pmatrix};$$

$$Q = \begin{pmatrix} 2C \sum_{i=1}^n \hat{x}_i \hat{x}_i^T + I & 2C \sum_{i=1}^n \hat{x}_i \\ 2C \sum_{i=1}^n \hat{x}_i & 2Cn \end{pmatrix}.$$

Управляющий параметр C подбирается экспериментально с помощью процедуры кросс-валидации [11, 13].

Полученные значения матрицы весов \bar{u} позволяют перейти от исходных *усиленных* признаков \hat{x}_i к *накопительным* признакам \hat{a}_i классифицируемых изображений:

$$\hat{a}_i = (\langle \bar{u}, \hat{x}_i \rangle + \bar{u}_0). \quad (2.7)$$

С учетом данного перехода к *накопительным* признакам задача восстановления глобальной регрессии по всему диапазону возрастов (y_{min} , y_{max}) при использовании кусочно-линейной ε -чувствительной функции потерь со значением чувствительности ε_0 формулируется следующим образом (см. параграф 1.7.3):

$$\begin{cases} \left\{ \frac{1}{2} \|\bar{w}\|^2 + C \sum_{i=1}^n (\xi_i^+ + \xi_i^-) \right\} \rightarrow \min_{w, w_0, \xi_i^+, \xi_i^-}; \\ y_i - \varepsilon_0 - \xi_i^- \leq (\langle \bar{w}, \hat{a}_i \rangle + w_0) \leq y_i + \varepsilon_0 + \xi_i^+; \\ \xi_i^+ \geq 0, \xi_i^- \geq 0, i = 1, n. \end{cases} \quad (2.8)$$

Отметим, что в соответствии с предложенной процедурой *усеченного* бутстраппинга в качестве обучающей выборки используется выборка \hat{G}_1^l .

Данная задача по-прежнему является задачей квадратичного программирования, которая решается стандартным образом методом множителей Лагранжа с использованием двойственных переменных λ_i^+ , λ_i^- . С учетом перехода к *накопительным* признакам итоговое соотношение для приближенной решающей функции, вычисляющей на основе глобальной регрессии возраст произвольного объекта по его вектору *накопительных* признаков \hat{a} , принимает вид:

$$f_0(\hat{a}, \varepsilon_0, \hat{G}^l) = \sum_{i=1}^n (\lambda_i^+ + \lambda_i^-) \langle \hat{a}, \hat{a}_i \rangle + w_0, \quad (2.9)$$

где параметр w_0 определяется из ограничений-неравенств, которые становятся равенствами на опорных векторах обучающей выборки \hat{G}_1^l .

$$f_0(\hat{a}, \varepsilon_0, \hat{G}^l) = (\langle \bar{w}, \hat{a}_i \rangle + w_0) = \begin{cases} y_i + \varepsilon_0, 0 < \lambda_i^+ < C, \lambda_i^- = 0, \xi_i^+ = \xi_i^- = 0; \\ y_i - \varepsilon_0, 0 < \lambda_i^- < C, \lambda_i^+ = 0, \xi_i^+ = \xi_i^- = 0. \end{cases}$$

Далее (в соответствии с предложенной процедурой *усеченного* бутстраппинга) производится обучающее тестирование полученной решающей функции на выборке \hat{G}_2^l . Затем выделяются «трудные» примеры – объекты, не попавшие в полосу чувствительности ε_0 , из которых удаляются два «самых трудных» примера с наибольшими значениями ошибки, и формируется новая обучающая выборка \hat{G}_{New}^l . По этой выборке повторяются рассмотренные вычисления, которые приводят к окончательному значению приближенной решающей функции $f_0(\hat{a}, \varepsilon_0, \hat{G}^l)$.

Абсолютно аналогичным образом решается задача локальной регрессии для каждого диапазона возрастов. Отличие состоит лишь в том, что вместо диапазона возрастов (y_{min}, y_{max}) используются диапазоны возрастов $(y - d^-(y), y + d^+(y))$, для каждого значения y , а вместо значения чувствительности ε_0 используются значения $\varepsilon(y)$. В результате формируется множество решающих функций $f_y(\hat{a}, \varepsilon(y), d^+(y), d^-(y), \hat{G}^l)$, позволяющих определить значение возраста для любого объекта по его *накопительным* признакам, вычисленным по *усиленным*. Отметим, что в процессе локальной регрессии используются те же *накопительные* признаки, которые были использованы при глобальной регрессии.

Вопрос исследования точности определения атрибута «возраст» по критериям *MAE* и *CS* с помощью предложенного алгоритма рассмотрен в главе 3.

2.5. Метод последовательного определения атрибутов личности по изображению лица

Предлагаемый в настоящей работе итоговый метод определения атрибутов личности основан на следующем утверждении: определение очередного атрибута личности можно осуществить с большей точностью (по выбранному критерию), если предыдущие атрибуты уже известны. В рассматриваемом случае это означает, что определить атрибут «раса» можно с большей точностью (по критерию *Accuracy*), если известна половая группа объекта, а определить атрибут «возраст» можно с большей точностью (по критериям *MAE* и *CS*), если известна половая группа и расовая принадлежность объекта. Исходя из данного утверждения, предлагается следующий итоговый алгоритм определения атрибутов личности, основанный на предложенных ранее подходах.

Состав обучающих выборок

Атрибут	Состав
«Пол»	Изображения мужчин и женщин разных рас и возрастов
«Раса»	Изображения мужчин разных рас и возрастов
	Изображения женщин разных рас и возрастов
Приближенное значение атрибута «возраст» (глобальная регрессия)	Изображения мужчин европеоидов разных возрастов
	Изображения женщин европеоидов разных возрастов
	Изображения мужчин монголоидов разных возрастов
	Изображения женщин монголоидов разных возрастов
	Изображения мужчин негроидов разных возрастов
	Изображения женщин негроидов разных возрастов
Уточненное значение атрибута «возраст» (локальная регрессия)	Изображения мужчин европеоидов в диапазоне возрастов $(y_1 - d_1^-, y_1 + d_1^+)$
	...
	Изображения мужчин европеоидов в диапазоне возрастов $(y_r - d_r^-, y_r + d_r^+)$
	Изображения женщин европеоидов в диапазоне возрастов $(y_1 - d_1^-, y_1 + d_1^+)$
	...
	Изображения женщин европеоидов в диапазоне возрастов $(y_r - d_r^-, y_r + d_r^+)$
	Изображения мужчин монголоидов в диапазоне возрастов $(y_1 - d_1^-, y_1 + d_1^+)$
	...
	...
	Изображения женщин монголоидов в диапазоне возрастов $(y_r - d_r^-, y_r + d_r^+)$
	Изображения женщин негроидов в диапазоне возрастов $(y_1 - d_1^-, y_1 + d_1^+)$
	...
	Изображения женщин негроидов в диапазоне возрастов $(y_r - d_r^-, y_r + d_r^+)$

1) Формирование *усиленных* признаков $\hat{\mathbf{X}}^l$ для обучающей выборки изображений (см. параграф 2.1):

- выделение половины нормализованного изображения лица и формирование по нему методом *LBP* вектора признаков, содержащего только *равномерные* шаблоны (см. параграф 1.7.1);
- использование метода *Adaboost* по атрибуту «пол» для исключения из вектора признаков изображения незначимых признаков (сумма весовых коэффициентов *усиленных* признаков равна 0,99).

2) Формирование решающей функции $f^{\text{пол}}(\hat{x})$ для бинарной классификации объектов методом *SVM* по атрибуту «пол» по *усиленным* признакам, вычисленным на первом шаге алгоритма, с учетом предложенной процедуры *усеченного* бутстраппинга (см. параграф 2.2).

3) Формирование решающих функции $f_{\text{м}}^{\text{раса}}(\hat{x})$ и $f_{\text{ж}}^{\text{раса}}(\hat{x})$ внутри каждой половой группы для множественной классификации объектов по атрибуту «раса» по *усиленным* признакам, сформированным на первом шаге алгоритма, с использованием бинарной классификации (схема - «один против всех») с учетом предложенной процедуры *усеченного* бутстраппинга (см. параграфы 1.6.2 и 2.3). Всего формируется по две решающих функции для каждой половой группы:

- функция $f_{\text{ж}}^{\text{Е}}(\hat{x})$ - выделяет среди объектов, отнесенных на втором шаге к классу «женщины», объекты, относящиеся к классу «европеоиды»;
- функция $f_{\text{ж}}^{\text{М}}(\hat{x})$ - выделяет среди объектов, отнесенных на втором шаге к классу «женщины», объекты, относящиеся к классу «монголоиды», а оставшиеся объекты - к классу «негроиды»;
- функция $f_{\text{м}}^{\text{Е}}(\hat{x})$ - выделяет среди объектов, отнесенных на втором шаге к классу «мужчины», объекты, относящиеся к классу «европеоиды»;

- функция $f_M^M(\hat{x})$ - выделяет среди объектов, отнесенных на втором шаге к классу «мужчины», объекты, относящиеся к классу «монголоиды», а оставшиеся объекты - к классу «негроиды».

4) Формирование совокупности решающих функций для двухэтапной схемы определения возраста методом **SVR** (см. параграф 2.3) для каждой расово-половой группы:

- формируются *накопительные* признаки \hat{A}^{lp} по *усиленным* признакам \hat{X}^{lp} (см. параграф 1.7.3) обучающей выборки для каждого сочетания половой («мужчины», «женщины») и расовой («европеиды», «монголоиды», «негроиды») группы p ; $p=(1, 6)$;
- по *накопительным* признакам \hat{A}^{lp} решается задача восстановления регрессии методом **SVR** по всей обучающей выборке расово-половой группы p с заданным значением ε -чувствительности ε_0^p , т.е. формируется решающая функция $f_0^p(\hat{a}^p, \varepsilon_0^p)$ (в общем случае, чувствительность ε функции потерь может зависеть от расово-половой группы p) для предварительного определения значение возраста (этап глобальной регрессии двухэтапной схемы определения возраста) для каждой расово-половой группы p , $p=(1, 6)$; в результате формируется шесть решающих функций, каждая из которых способна для любого объекта (изображения лица) с определенными на предыдущих шагах атрибутами «пол» и «раса» вычислить приближенное значение атрибута «возраст» по *усиленным* признакам этого объекта; при формировании решающих функций $f_0^p(\hat{a}^p, \varepsilon_0^p)$ по каждой расово-половой группе p применяется процедура *усеченного* бутстраппинга на всей обучающей выборке;
- для каждого значения возраста y^p в каждой расово-половой группе p выделяется диапазон возрастов $(y^p - d^{p-}(y^p), y^p + d^{p+}(y^p))$ с

значениями $d^{p-}(y^p)$ и $d^{p+}(y^p)$ (в общем случае, ширина диапазона может зависеть от расово-половой группы p), которые определяются экспертом исходя из специфики решаемой задачи;

- внутри каждого диапазона, относящегося к значению y^p , по соответствующему подмножеству прецедентов (x_i^p, y_i^p) : $y_i^p \in (y^p - d^{p-}(y^p), y^p + d^{p+}(y^p))$; $p = (1, 6)$; $i = (1, n^p)$; n^p - число прецедентов обучающей выборки в расово-половой группе p , решается задача восстановления регрессии методом **SVR** (с учетом перехода к *накопительным* признакам), в которой чувствительность функции потерь $\varepsilon^p(y^p)$ зависит от значения y^p и так же определяется экспертом, исходя из специфики решаемой задачи (в общем случае, чувствительность ε функции потерь может зависеть от расово-половой группы p); в результате формируется совокупность решающих функций $f_y^p[\hat{a}^p, \varepsilon^p(y^p), d^{p+}(y^p), d^{p-}(y^p)]$ для определения уточненного значения возраста для каждой расово-половой группы p и возраста y^p : $p=(1, 6)$, $y^p \in (y_{min}, y_{max})$; при формировании решающей функции $f_y^p[\hat{a}^p, \varepsilon^p(y^p), d^{p+}(y^p), d^{p-}(y^p)]$ по каждой расово-половой группе p в каждом диапазоне возрастов $(y^p - d^{p-}(y^p), y^p + d^{p+}(y^p))$ применяется процедура *усеченного* бутстраппинга.

Дерево решений предлагаемого метода приведено на рис. 2.5.

Применение процедуры *усеченного* бутстраппинга в рамках метода последовательного определения атрибутов личности позволяет удалить из обучающей выборки 22 «очень плохих» прецедента: два на этапе половой классификации, еще 8 на этапе расовой классификации, и, наконец, еще 12 на этапе глобальной регрессии при определении возраста.

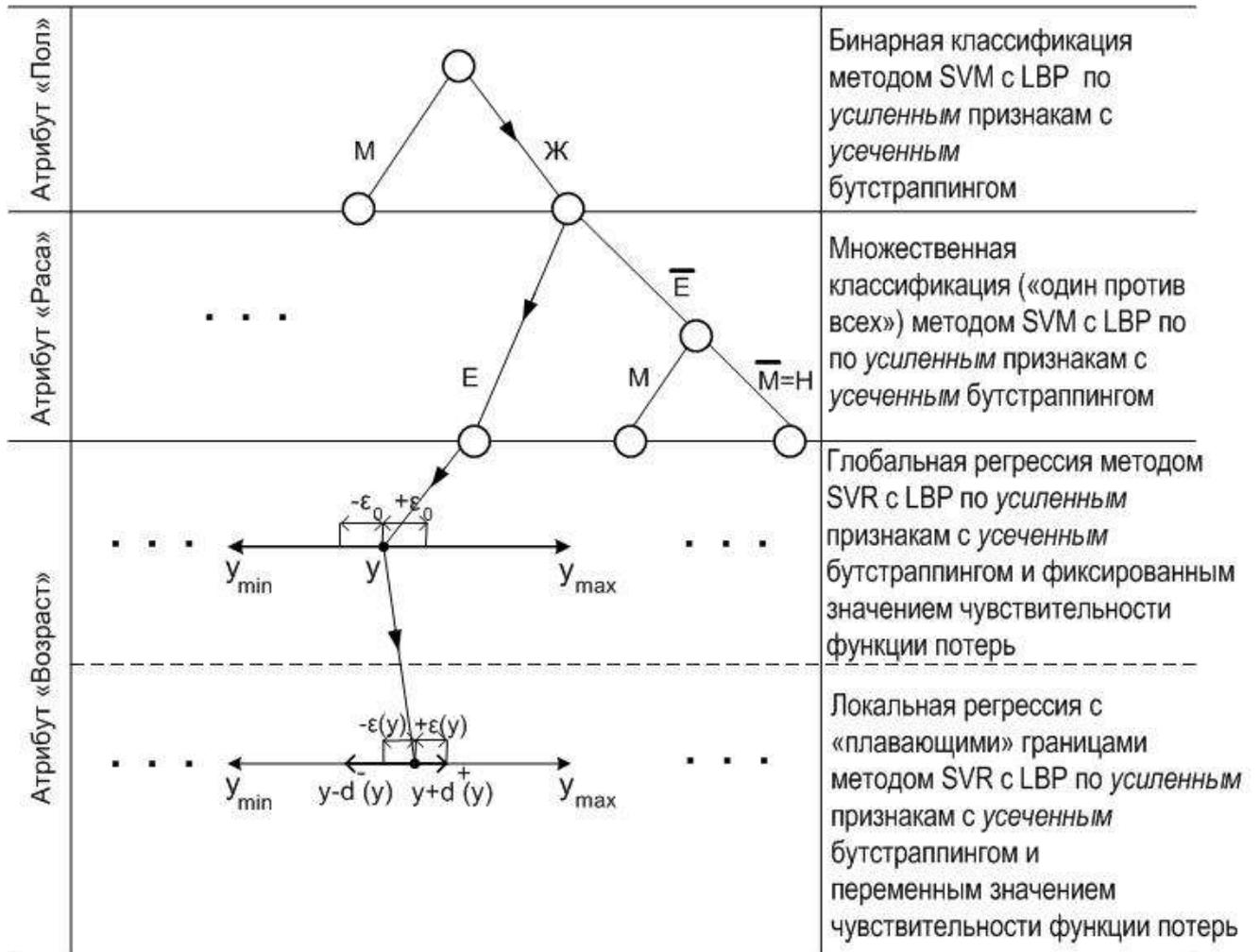


Рис. 2.5. Дерево решений метода последовательного определения атрибутов личности по изображению лица

Вновь отметим важную особенность предложенного метода последовательного определения атрибутов – его инвариантность по отношению к используемым на каждом шаге методам формирования решающих функций.

2.6. Выводы по главе 2

1) В отличие от существующего эмпирического описания локальных бинарных шаблонов и *равномерных* локальных бинарных шаблонов,

используемых для признакового описания изображений, предложено формализованное описание указанных признаков на базе принципа конечной топологии, что является предпосылкой для унифицированного функционального представления изображений.

2) Разработан метод снижения размерности пространства признаков изображений, включающий автоматическую процедуру определения значимости каждого признака на базе алгоритма *Adaboost*, который позволил, за счет использования априорной информации о человеческом лице, снизить размерность пространства признаков изображения почти в 10 раз по сравнению со стандартным методом *LBP*.

3) Предложена модификация метода бинарной классификации, отличающаяся использованием применявшейся ранее при решении других задач компьютерного зрения процедуры бутстраппинга (обучение на «трудных» примерах). Данная модификация используется для классификации изображений по атрибуту «пол» и атрибуту «раса».

3) Процедура бутстраппинга, усовершенствована путем исключения «самых трудных» примеров (*усеченный бутстраппинг*) для снижения эффекта переобучения решающей функции с целью снижения влияния на ее формирование незначимого для классификации «шума» в обучающей выборке.

4) Предложена модификация метода множественной классификации как последовательности бинарных классификаций, отличающаяся использованием предложенной модифицированной *усеченным бутстраппингом* бинарной классификации.

5) Предложена модификация двухэтапной схемы определения атрибута «возраст» методом восстановления регрессии на основе опорных векторов (*SVR*), отличающаяся использованием этапе локальной регрессии диапазону возрастов с «плавающими» границами, значение которых зависит от возраста. При этом в ε -чувствительной функции

потерь чувствительность ε также зависит от значения возраста, а не является постоянной. В процессе формирования решающих функций используется процедура *усеченного бутстраппинга*.

5) Разработан метод последовательного определения атрибутов личности («пол» → «раса» → приближенное значение атрибута «возраст» → уточненное значение атрибута «возраст»), отличающийся учетом на каждом следующем этапе результаты предыдущего и использующий перечисленные выше модификации существующих методов для определения атрибутов личности.

6) Предложенные подходы могут быть применены и к другим методам формирования пространства признаков и методам построения решающих функций для повышения качества классификации и не только для решения задачи определения атрибутов личности.

3. ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ РАЗРАБОТАННОГО МЕТОДА ПОСЛЕДОВАТЕЛЬНОГО ОПРЕДЕЛЕНИЯ АТТРИБУТОВ ЛИЧНОСТИ ПО ИЗОБРАЖЕНИЮ ЛИЦА

3.1. Методика исследования точности определения атрибутов

Вопрос о качестве алгоритма обучения (решающей функции), синтезированного по конечным выборкам прецедентов, является фундаментальной проблемой теории машинного обучения (*machine learning theory*) [16]. В процессе формирования решающей функции решается задача оптимизации с целью подбора таких параметров, при которых решающая функция допускает наименьшее число ошибок на обучающей выборке, т.е. осуществляется некоторая «подгонка» модели под выборку. Такой подход получил название *минимизации эмпирического риска* [16]. На практике исследователи столкнулись с проблемой *переобучения*: решающая функция слишком хорошо подстраивалась под конкретный набор прецедентов, а за пределами обучающей выборки *обобщающая способность* сформированной решающей функции снижалась [11, 13]. Из данного наблюдения был сделан вывод о существовании оптимальной сложности модели, при которой достигается наилучшее качество обобщения. Первое формальное обоснование данного практического опыта было дано в работе [9], а в настоящее время развито в работах [15, 16] в рамках комбинаторного подхода к оценке качества обучаемых алгоритмов.

Основным постулатом комбинаторного подхода является утверждение: метод скользящего контроля (*cross validation*) [73] адекватно оценивает качество алгоритма обучения по прецедентам [14, 15, 16]. Поэтому, применительно к решаемой в настоящей работе задаче

определения атрибутов личности по изображению лица на основе методов *SVM/SVR* для проведения экспериментальных исследований качества предлагаемого подхода для формирования решающей функции целесообразно использовать метод скользящего контроля [73].

Процедура скользящего контроля заключается в следующем. Фиксируется некоторое множество разбиений исходной выборки прецедентов G , $|G| = m$, состоящей из пар прецедентов, на две подвыборки: обучающую G^l и тестовую G^t :

$$G^l \cap G^t = \emptyset.$$

Для каждого разбиения проводится формирование решающей функции f по обучающей подвыборке G^l с последующей оценкой ее качества (по выбранному критерию качества Q) на тестовой подвыборке G^t . Оценка значения выбранного критерия качества Q методом скользящего контроля определяется как среднее по всем разбиениям. Существует несколько различных способов разбиения исходной выборки G на подвыборки, каждый из которых приводит к некоторой разновидности метода [73].

Для исследования качества предлагаемого в настоящей работе подхода предлагается использовать наиболее распространенный на сегодня метод k -кратного скользящего контроля [16]. Данная модификация предполагает разбиение исходной выборки G случайным образом на k непересекающихся подвыборок $g_1, g_2 \dots g_k$, из которых каждый раз $(k-1)$ подвыборка используется для обучения, а оставшаяся – для контроля, т.е.

$$G_1^l = (g_1 \cup g_2 \dots \cup g_{k-1})$$

$$G_1^t = g_k$$

$$G_2^l = (g_1 \cup g_2 \dots \cup g_{k-2} \cup g_k)$$

$$G_2^t = g_{k-1}$$

...

$$G_k^l = (g_2 \cup g_3 \dots \cup g_k)$$

$$G_k^t = g_1$$

При этом для каждого разбиения (G_i^l, G_i^t) вычисляется выбранный функционал качества Q_i , а итоговая оценка данного функционала вычисляется как среднее \bar{Q} по всем разбиениям.

В этом случае, полученное среднее значение функционала качества \bar{Q} является несмещенной оценкой данного функционала [16]. Традиционным является разбиение на 5 - 10 подвыборок [15, 16], т.е. каждый раз от 80% до 90% исходной выборки используется для обучения, а соответственно от 20% до 10% - для тестирования. В итоге, все доступные прецеденты участвуют как в процессе обучения, так и в процессе тестирования.

Необходимо отметить, что в настоящей работе для выполнения приведенных далее в настоящей главе исследований количество подвыборок (5 или 10) выбиралось исходя из размерности соответствующей исходной выборки G , с целью обеспечить участие в каждом эксперименте нескольких тысяч прецедентов. При этом для каждого разбиения выборки G в процессе обучения использовалась предложенная процедура усеченного бутстраппинга (см. параграф 2.2) с целью исключения нетипичных прецедентов.

3.2. Исследуемые наборы данных

3.2.1. Существующие наборы данных с изображениями лиц

Обучение классификаторов для определения атрибутов личности по изображению лица предполагает наличие базы таких изображений с размеченными значениями исследуемых атрибутов. В открытом доступе существует не так много баз изображений лиц с размеченным полом,

расой и возрастом. При этом при наличии базы изображений лиц без отметок о поле или расе представляется возможным вручную обработать такую базу и указать соответствующие значения признаков. В случае с атрибутом «возраст» провести аналогичную процедуру практически невозможно, так как даже человек не может по изображению лица определить значение возраста.

В приведенной ниже таблице указаны доступные базы изображений лиц людей с указанием их параметров:

Таблица 3.1.

Параметры открытых баз изображений лиц людей

№ п/п	Наименов. базы	Кол-во изображ.	Атрибут «пол»	Атрибут «раса»	Атрибут «возраст»	Примечание
1	FERET [58]	1755	+			
2	MORPH [94]	55134	+		+	
3	FG-NET [59]	1002	+		+	
4	Mall [90]		+	+		
5	LFW [78]	13233	+	+	+	от 20 до 70 лет

Анализ доступных баз изображений лиц показал, что для решения поставленной задачи исследования точности последовательного определения атрибутов «пол», «раса» и «возраст» указанные базы могут быть использованы лишь ограниченно (можно исследовать точность определения отдельных атрибутов). Необходимо также отметить, что большая часть баз содержат изображения, полученные при специальной съемке людей, что приводит к завышенным значениям качества классификации по сравнению с анализом изображений, полученным в реальных условиях. Наиболее подходящей для целей настоящего исследования является база LFW, но она практически не содержит изображений лиц до 20 лет, а возрастная классификация именно в этом диапазоне в ряде случаев представляется наиболее практически важной

(см. параграф 1.2). В связи с этим в настоящей работе поставлена и решена задача формирования собственной базы изображений лиц по открытым источникам (социальным сетям) в сети Интернет.

3.2.2. Формирование набора данных изображений лиц по открытым источникам в сети Интернет

Благодаря широкому развитию социальных сетей (Facebook, Одноклассники, ВКонтакте и пр.) в сети Интернете появилось много фотографий людей, для которых известны или могут быть определены искомые атрибуты: «пол», «раса», «возраст». Задача формирования базы изображений лиц большого размера, размеченной по данным атрибутам, существенно упрощается благодаря наличию специальных инструментов - сервис-агрегаторов данных из социальных сетей, которые позволяют производить поиск по всем доступным социальным сетям через единый интерфейс (<http://people.yandex.ru>). В процессе выполнения настоящей работы было разработано соответствующее программное обеспечение для автоматической загрузки фотографий из открытых источников сети Интернет, которые затем использовались для создания базы изображений лиц, названной BigSample. Главным критерием для отбора изображений было наличие информации о возрасте человека. Кроме того, с помощью простейших классификаторов по полу и расе выполнялась предварительная разметка загружаемых изображений по этим атрибутам. Далее, все полученные изображения просматривались вручную с целью проверки и уточнения значений атрибута «пол» и атрибута «раса». В итоге была сформирована база, содержащая в себе 169629 изображений лиц. Примеры изображений приведены на Рис. 3.1.

Расовый и половой состав базы изображений BigSample приведен в Табл. 3.2, а распределение изображений по возрасту – на Рис. 3.2.



а)

б)

в)

Рис. 3.1. Примеры изображений базы BigSample

а) женщина, негроид, 22 года

б) мужчина, монголоид, 69 лет

в) мужчина, европеид, 21 год

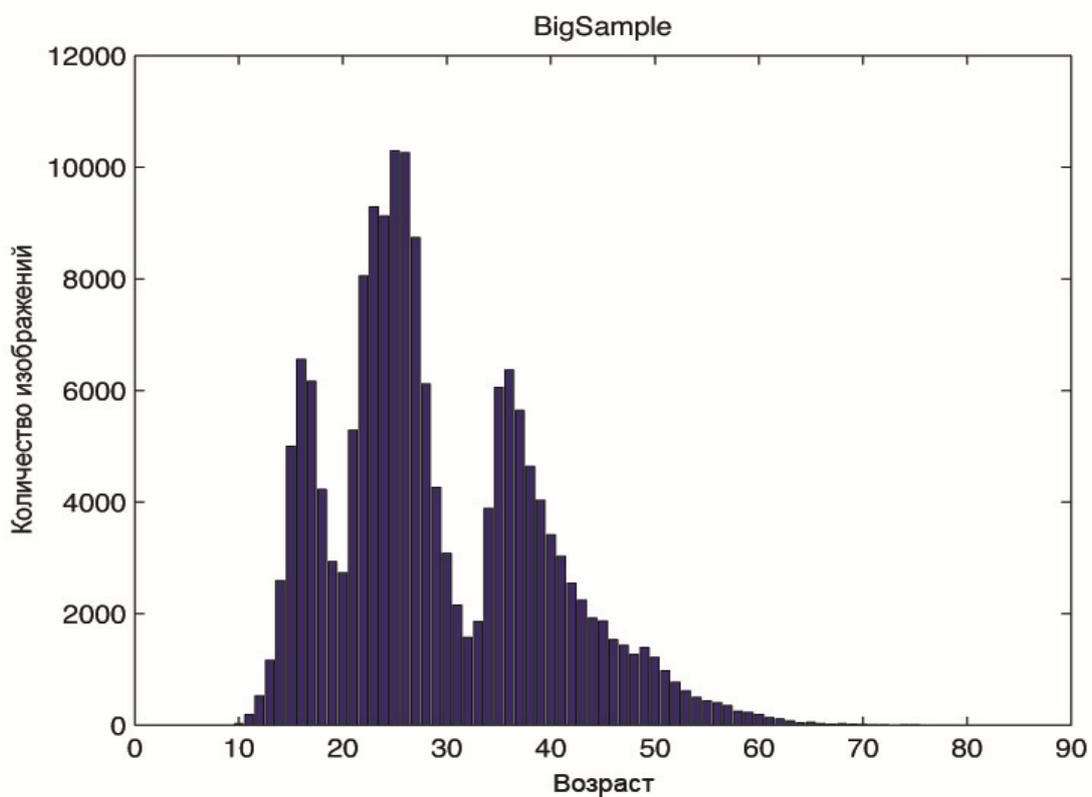


Рис. 3.2. Распределение количества изображений людей по возрасту в базе BigSample

Таблица 3.2.

Расовый и половой состав людей в базе BigSample

	<i>Мужчины</i>	<i>Женщины</i>	Всего изображений
<i>Европеоиды</i>	50190	42429	92619
<i>Монголоиды</i>	22740	19682	42422
<i>Негроиды</i>	18345	16533	34878
Всего изображений	91275	78 654	169629

Отметим, что примерно через год после завершения формирования автором указанной базы изображений и публикации некоторых результатов исследований на ее основе [121] коллективом авторов аналогичным способом по открытым источникам в сети Internet была сформирована база, содержащая около 500 000 изображений (<https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/>). К сожалению, в этой базе нет отметок о расе человека. При половой классификации по этой базе, очевидно, будут получены данные по качеству классификации аналогичные данным, полученным по авторской базе.

3.3. Результаты исследования точности разработанных модификаций и метода последовательного определения атрибутов личности

3.3.1. Точность определения атрибута «пол»

Исследование точности определения атрибута «пол» по изображению лица осуществлялось по доступным открытым базам изображений FERET [58] и MORPH [94], и по сформированной в рамках данной работы базе изображений BigSample (см. параграф 3.2.2). При этом использовалась методика скользящего контроля (см. параграф 3.1), в

соответствии с которой прецеденты в базах FERET и MORPH из-за малого количества изображений делились на 5 частей (5-кратный скользящий контроль), а в базе BigSample – на 10 частей (10-кратный скользящий контроль). Поскольку количество изображений мужчин и женщин в исследуемых наборах данных было примерно равным, использовалась метрика *Accuracy* (см. параграф 1.6). Результаты исследования приведены в Табл. 3.3.

По результатам проведенного исследования можно сделать следующие выводы:

Таблица 3.3

Результаты исследования точности определения атрибута «пол»

Метод		<i>SVM</i>	<i>SVM</i>	<i>SVM</i> + «усеченный» бутстраппинг	Повышение качества
Признаки		<i>LBP</i>	«усиленные» <i>LBP</i>	«усиленные» <i>LBP</i>	
<i>Accuracy</i>	FERET	86,7%	87,4%	94,4%	7,7%
	MORPH	87,0%	87,8%	95,2%	8,2%
	BigSample	78,5%	79,4%	92,1%	13,6%

- 1) Точность определения атрибута «пол» известным методом *SVM* на основе *LBP* на базе изображений BigSamle практически совпадает с опубликованными результатами. полученными на реальных изображениях из сети Интернет (в [92] при использовании метода *SVM* на базе *LBP* получена точность классификации по критерию *Accuracy* 76,71%). Несколько более высокая точность определения атрибута «пол» по базам FERET и MORTH объясняется использованием в них, в отличие от базы BigSample, специально подготовленных изображений.

- 2) Переход к «усиленным» признакам *LBP* (см. параграф 2.1) приводит к повышению точности определения атрибута «пол» как для специально подготовленных баз изображений лиц людей (базы FERET и MORPH), так и для реальных изображений из сети Интернет (база BigSample). Следовательно, использование «усиленных» признаков представляется обоснованным и целесообразным.
- 3) Применение процедуры «усеченного» бутстраппинга приводит к повышению точности определения атрибута «пол» как для специально подготовленных баз изображений лиц людей, так и для реальных изображений из сети Интернет примерно на 13% по сравнению с базовым методом. Следовательно, применение данной процедуры представляется обоснованным и целесообразным.

3.3.2. Точность последовательного определения атрибута «раса» после определения атрибута «пол»

Исследование точности определения атрибута «раса» по изображению лица осуществлялось по доступным открытой базе изображений MALL [90] и по сформированной в рамках данной работы базе изображений BigSample (см. параграф 3.2.2). При этом, как и при исследовании точности определения атрибута «пол», использовалась методика скользящего контроля (см. параграф 3.1), в соответствии с которой прецеденты в базах PAL и MALL из-за малого количества изображений делились на 5 частей, а в базе BigSample – на 10 частей. Поскольку в базе Количество изображений людей разных рас из базы BigSample отбиралось таким образом, чтобы сохранить естественное соотношение между расами (2:1,5:1). Такой подбор позволил использовать

для оценки точности классификации меру *Accuracy*. Отметим, что в соответствии с предложенным методом последовательного определения атрибутов личности, главное внимание уделялось точности определения атрибута «раса» внутри выделенной половой группы. Результаты исследования приведены в Табл. 3.4.

Таблица 3.4

Результаты исследования точности определения атрибута «раса»

Метод		<i>SVM</i>	<i>SVM</i> «усечен- ный» бутстрап.	<i>SVM</i> + «усеченный» бутстрап. в заданной половой группе	Опред. атрибута «раса» после атрибута «пол»	Повыше- ние качества
Признаки		<i>LBP</i>	«усил.» <i>LBP</i>	«усил.» <i>LBP</i>	«усил.» <i>LBP</i>	
<i>Accuracy</i>	Mall	81,3%	90,4%	94,3%	93,5%	12,2%
	BigSampe	74,1%	84,9%	91,1%	89,2%	15,1%

По результатам проведенного исследования можно сделать следующие выводы:

- 1) Использование «усиленных» признаков *LBP* и процедуры «усеченного» бутстраппинга приводит к повышению точности определения атрибута «раса» как для специально подготовленных баз изображений лиц людей, так и для реальных изображений из сети Интернет примерно на 9% и 11% соответственно. Следовательно, использование «усиленных» признаков *LBP* и процедуры «усеченного» бутстраппинга для определения атрибута «раса» представляется обоснованным и целесообразным.
- 2) Определение атрибута «раса» внутри заданной половой группы приводит к повышению точности определения атрибута «раса» еще примерно на 4% и 6% для этих же баз. Следовательно, использование «усиленных» признаков и процедуры «усеченного»

бутстраппинга внутри выделенной половой группы для определения атрибута «раса» представляется обоснованным и целесообразным.

- 3) Совместное использование «усиленных» признаков, процедуры «усеченного» бутстраппинга и последовательное определение атрибута «раса» после определения атрибута «пол» приводит к повышению точности определения атрибута «раса» по изображению на 12% и 15% для указанных баз соответственно по сравнению с базовым алгоритмом.

3.3.3. Точность определения атрибута «возраст» после определения атрибутов «пол» и «раса»

Исследование точности определения значения атрибута «возраст» по изображению лица осуществлялось по доступным открытым базам изображений FG-NET [59] и MORPH [94], и по сформированной в рамках данной работы базе изображений BigSample (см. параграф 3.2.2). При этом, как и в предыдущем параграфе, использовалась методика скользящего контроля (см. параграф 3.1), в соответствии с которой прецеденты в базах FG-NET и MORPH из-за малого количества изображений делились на 5 частей, а в базе BigSample – на 10 частей. В соответствии с предложенным методом последовательного определения атрибутов личности точность определения значения атрибута «возраст» оценивалась для выделенной на предыдущих этапах расово-половой группы. Для оценки точности использовались метрики *MAE* и *CS(e)* (см. параграф 1.6). Результаты проведенного исследования приведены в Табл. 3.5.

По результатам проведенного исследования можно сделать следующие выводы:

1) Использование предложенной модификации двухэтапной схемы восстановления регрессии для определения атрибута «возраст» (см. параграф 2.3) совместно с использованием *усиленных* признаков описания изображения (см. параграф 2.2) приводит к повышению точности определения данного атрибута на реальных изображениях лиц людей примерно на 0,22, 0,19 и 1,35 года по критерию *MAE*, на 7%, 7% и 5% по критерию *CS(3)* и на 9%, 8% и 10% по критерию *CS(5)* для трех указанных баз соответственно. Следовательно, использование данной модификации представляется целесообразным.

Таблица 3.5

Результаты исследования точности определения атрибута «возраст»

Метод		<i>SVR</i> (два этапа)	<i>SVR</i> (два этапа) + «плавающие» границы + «усеченный» бутстрап.	<i>SVR</i> (два этапа) + «плавающие» границы + «усеченный» бутстрап. внутри заданной расово- половой группы	Определ. атрибута «возраст» после атрибутов «пол» и «раса»	Повы- шение качества
Признаки		<i>LBP</i>	«усил.» <i>LBP</i>	«усил.» <i>LBP</i>	«усил.» <i>LBP</i>	
<i>MAE</i>	FG-NET	4,65	4,43	4,30	4,31	0,34
	MORPH	4,61	4,42	4,28	4,29	0,32
	BigSample	8,07	6,72	6,08	6,09	1,98
<i>CS(3)</i>	FG-NET	38%	45%	51%	49%	11%
	MORPH	37%	44%	49%	48%	11%
	BigSample	21%	26%	42%	41%	20%
<i>CS(5)</i>	FG-NET	53%	62%	72%	71%	18%
	MORPH	52%	60%	72%	71%	19%
	BigSample	42%	52%	63%	62%	20%

- 2) Определение атрибута «возраст» с помощью предложенной модификации двухэтапной схемы восстановления регрессии внутри выделенной расово-половой группы повышает точность определения данного атрибута по критерию *MAE* на 0,35, 0,33 и 1,99 года, на 13%, 12% и 21% по критерию *CS(3)* и на 19%, 20% и 21% по критерию *CS(5)* для трех рассмотренных баз соответственно. Следовательно, формирование решающей функции для каждой расово-половой группы является продуктивным.
- 3) Некоторое снижение точности определения атрибута «возраст» в процессе реализации предложенного метода последовательного определения атрибутов («пол» «раса» «возраст») объясняется тем, что классификатор иногда допускает ошибки в определении пола и расы. Это приводит к тому, что на этапе определения возраста применяется решающая функция, обученная на «чужой» расово-половой группе. Тем не менее, итоговый результат – повышение точности определения атрибута «возраст» по критерию *MAE* на 0,34, 0,32 и 1,98 года, на 11%, 11% и 20% по критерию *CS(3)* и на 18%, 19% и 20% по критерию *CS(5)* на рассмотренных базах изображений лиц людей, позволяет рекомендовать предложенный метод для использования в системах компьютерного зрения.

3.4. Исследование затрат времени на определение атрибутов личности по изображению лица разработанным методом

Центральной идеей снижения затрат времени на процедуру обучения и последующую процедуру классификации изображений лиц является снижение размерности пространства *LBP*-признаков, основанной на использовании априорной информации об изображении лица и формировании «усиленных» признаков (см. параграф 2.2). Результаты

численных экспериментов по выполнению наиболее трудоемкой процедуры – определения возраста, проведенные на существующей открытой базе изображений лиц MORTH [94] и собственной базе изображений BigSample (см. параграф 3.2.1) показали, что предлагаемый подход, даже при значении $\delta = 0,99$ (оставляем признаки, имеющие хоть какую-нибудь значимость для классификации), позволяет отбросить более половины признаков и сократить размерность пространства признаков примерно до 480, т.е. почти в 5 раз по сравнению с традиционным использованием равномерных *LBP*, а для порога значимости $\delta = 0,95$ число *усиленных* признаков сокращается до 280, что почти на порядок меньше исходного.

В Табл. 3.6 приведено сравнение результатов, полученных в [120], с результатами, полученными аналогичным способом на идентичном компьютере, но только при использовании *усиленных LBP*, сформированных предлагаемым в настоящей работе методом. Отметим, что в работе [120], для сокращения времени классификации предлагалось использовать ряд упрощений в базовом алгоритме, что приводило к снижению точности классификации. Использование *усиленных* признаков позволило добиться аналогичного сокращения времени классификации без потери точности.

Таблица 3.6.

Сравнение результатов затрат времени на возрастную классификацию на основе *LBP* и *усиленных LBP*
(Процессор Intel Core i7-6700HQ, 2,4 GHz)

	Равномерные <i>LBP</i> N=2436 (58 x 42)	Половина изображения N=1218 (58 x 21)	<i>Усиленные</i> <i>LBP</i> N=480 ($\delta=0,99$)	<i>Усиленные</i> <i>LBP</i> N=280 ($\delta=0,95$)
<i>MAE</i>	4,52 [120]	4,53	4,37	4,45
Время класси- фикации, с	3,21 [120]	1,63	0,69	0,42

3.5. Программная реализация разработанного метода последовательного определения атрибутов личности по изображению лица

Программная реализация разработанных модификаций и метода последовательного определения атрибутов личности по изображению лица выполнена на языке программирования PYTHON. Выбор этого языка обусловлен простотой интеграции разработанных на нем приложений с существующими web-приложениями и наличием открытых библиотек, реализующих базовые алгоритмы *Adaboost* [114] и *SVM/SVR* [115], используемые в настоящей работе.

В настоящее время доступна версия 3.20 библиотеки LIBSVM (<https://www.csie.ntu.edu.tw/~cjlin/libsvm>). В составе библиотеки, наряду с модулями, реализующими бинарную и множественную классификацию на основе метода опорных векторов (C-SVC, nu-SVC), включен модуль восстановления регрессии на основе опорных векторов с использованием ϵ -чувствительной функции потерь (epsilon-SVR).

3.6. Выводы по главе 3

1) Анализ доступных открытых баз изображений лиц показал, что на их основе невозможно провести адекватное исследование качества разработанных подходов к определению атрибутов личности по изображению лица, так как эти базы содержат отметки либо только о некоторых атрибутах, либо имеют недостаточное для проведения исследований количество объектов. Поэтому в работе поставлена и решена задача создания собственной базы изображений, лишенной

указанных недостатков. На основе открытых источников информации в сети Интернет была сформирована база BigSample, содержащая 169 600 изображений мужчин и женщин трех рас в возрасте от 10 до 70 лет.

2) Разработанные модификации методов половой, расовой и возрастной классификации людей по изображению лица, объединенные в метод последовательного определения атрибутов личности, позволили на открытых базах изображений и сформированной в рамках данной работы собственной базе изображений лиц людей повысить качество распознавания (по критерию *Accuracy*) пола – на 13%, расы – на 15%, возраста по критерию *MAE* на 1,98 года и по критерию *CS(5)* – на 20% по сравнению с известными подходами.

3) Разработанный метод снижения размерности пространства признаков позволил сократить время обучения классификаторов и последующей совместной классификации пола, расы и возраста почти в 8 раз по сравнению с традиционными подходами с повышением качества определения атрибутов личности по используемым критериям.

ЗАКЛЮЧЕНИЕ

Основные результаты работы:

1) Анализ существующих методов определения атрибутов личности по изображению лица показал, что одним из перспективных для модификации является метод опорных векторов (*SVM*) при определении атрибутов «пол» и «раса», и регрессия на основе опорных векторов (*SVR*) при определении атрибута «возраст», использующие метод локальных бинарных шаблонов (*LBP*) для формирования вектора признаков изображения. При этом, из-за очень большой размерности формируемого методом *LBP* пространства признаков, возникает задача понижения размерности данного пространства с целью удаления незначущих признаков для повышения качества классификации и сокращения затрат времени как на обучение классификатора, так и на саму классификацию.

2) Разработан метод снижения размерности пространства признаков изображений, основанный на автоматической процедуре определения значимости каждого признака на базе алгоритма *Adaboost*. Данный метод позволил, за счет использования априорной информации о человеческом лице, снизить размерность пространства признаков изображения почти в 10 раз по сравнению со стандартным методом *LBP*, использующим *равномерные* бинарные шаблоны.

3) Предложена модификация метода бинарной классификации и основанной на бинарной – множественной классификации для определения атрибутов «пол» и «раса», отличающаяся использованием применявшейся ранее при решении других задач компьютерного зрения процедуры бутстраппинга (обучение на «трудных» примерах).

4) Предложено усовершенствование процедуры бутстраппинга, заключающееся в исключении «самых трудных» примеров (*усеченный* бутстраппинг) для снижения эффекта переобучения решающей функции с

целью снижения влияния на ее формирование незначимого для классификации «шума» в обучающей выборке.

5) Предложена модификация двухэтапной схемы определения атрибута «возраст» методом восстановления регрессии, отличающаяся тем, что:

- на этапе локальной регрессии используются диапазоны возрастов с «плавающими» границами, значение которых зависит от значения возраста, определенного на этапе глобальной регрессии;
- чувствительность ε в ε -чувствительной функции потерь также зависит от значения возраста, а не является постоянной;
- в процессе формирования решающих функций используется процедура *усеченного бутстраппинга*.

5) Разработан метод последовательного определения атрибутов личности («пол» → «раса» → приближенное значение атрибута «возраст» → уточненное значение атрибута «возраст»), отличающийся учетом на каждом следующем этапе результаты предыдущего и использующий перечисленные выше модификации существующих методов для определения атрибутов личности.

б) Разработанный метод последовательного определения атрибутов личности по изображению лица на существующих открытых базах изображений лиц людей и собственной базе изображений, сформированной по открытым источникам в сети Internet, позволил повысить качество распознавания (по критерию *Accuracy*) пола – на 13%, расы – на 15%, возраста по критерию *MAE* на 1,98 года и критерию *CS(5)* – на 20%. На базе реальных изображений лиц людей, сформированной по открытым источникам в сети Интернет, достигнута точность по критерию *Accuracy* определения атрибута «пол» - 92%, атрибута «раса» – 89% и атрибута «возраст» по критерию *MAE* – 6,09 лет.

7) Разработанный метод снижения размерности пространства признаков совместно с методом последовательного определения атрибутов личности позволили сократить время обучения классификаторов и совместной классификации пола, расы и возраста в 8 с повышением качества определения искомых атрибутов личности по используемым критериям.

Дальнейшее развитие методов определения атрибутов личности по изображению лица можно осуществлять в направлении:

- определения других атрибутов личности (например, выражение эмоций);
- поиска скрытых структурных особенностей в изображении лица человека с целью формирования методом *LBP* более информативного пространства признаков за счет использования принципа конечной топологии;
- разработки алгоритмов параллельной классификации за счет использования возможностей современных многоядерных и многопоточковых процессоров;
- использования предложенных подходов совместно с другими методами формирования решающих функций для определения атрибутов личности при решении других задач классификации изображений.

Литература

1. Амосов А.А., Дубинский Ю.А., Копченова Н.В. Вычислительные методы для инженеров. М.: Изд-во МЭИ, 2003. – 704 с.
2. Барлыбаев А.Б. Адаптивный интерфейс пользователей. Труды Международной научно-технической конференции OSTIS'14, 2014, с. 553-558.
3. Бонгард М. М. Проблема узнавания. М.: Наука, 1967. - 320 с.
4. Бредихин Р.Н. Об одном подходе к распознаванию оптических образов текстов // Вестник МЭИ, 2005, № 2, с. 134-141
5. Бредихин Р.Н. О принципе конечной топологии и его применении. ВИНТИ, №637 - В2004 (Мос-ковский энергетический институт (технический университет)), 2004. – 18 с.
6. Бунак В.В. Род Ното, его возникновение и последующая эволюция. М.: Наука, 1980. - 328 с.
7. Вагин В.П., Головина Е.Ю., Загорянская А.А., Фомина М.В. Достоверный и правдоподобный вывод в интеллектуальных системах. М.: ФИЗМАТЛИТ, 2004. – 704 с.
8. Вапник В. Н. Восстановление зависимостей по эмпирическим данным. М.: Наука, 1979. – 449 с.
9. Вапник В.Н., Червоненкис А.Я. Теория распознавания образов (статистические проблемы обучения). М.: Наука, 1974. – 415 с.
10. Визильтер Ю.В., Желтов С.Ю., Бондаренко А.В. и др., Обработка и анализ изображений в задачах машинного зрения. – М.: Физматкнига, 2010. – 672 с.
11. Вьюгин В.В. Математические основы теории машинного обучения и прогнозирования. М.: МЦМНО, 2013. – 390 с.
12. Возраст сексуального согласия. Википедия. Режим доступа: https://ru.wikipedia.org/wiki/Возраст_сексуального_согласия.

13. Воронцов К.В. Машинное обучение. Курс лекций. Электронный ресурс (<http://machinelearning.ru>)
14. Воронцов К.В. Комбинаторные оценки качества обучения по прецедентам // Докл.РАН. – 2004. – Т. 394, № 2. – С. 175-178
15. Воронцов К.В. Комбинаторный подход к оценке качества обучаемых алгоритмов // Математические вопросы кибернетики. – 2004. – Т. 394, № 2. – С. 175-178
16. Воронцов К.В. Обзор современных исследований по проблеме качества обучения алгоритмов. [Электронный ресурс]. – Режим доступа: <http://www.ccas.ru/frc/papers/voron04twim.pdf>
17. Гулин В.В. Сравнительный анализ методов классификации текстовых документов. Вестник МЭИ., №4. 2011. С.100-108.
18. Гулин В.В., Фролов А.Б. О классификации текстовых документов с учётом некоторых структурных особенностей // Известия РАН. Теория и системы управления, 2016, № 3, с. 66-75.
19. Дуда Р., Харт П. Распознавание образов и анализ сцен. М.: Мир, 1976. – 511 с.
20. Журавлев Ю.И. Избранные научные труды. М.: Издательство Магистр, 1998. – 420 с.
21. Журавлев Ю.И., Об алгебраическом подходе к решению задач распознавания или классификации. Проблемы кибернетики. М.: Наука, 1978, вып. 33, с. 5-68.
22. Журавлев Ю.И., Гуревич И.Б., Распознавание образов и распознавание изображений // Распознавание, классификация, прогноз, 1989, т. 2, с. 5-72.
23. Журавлев Ю.И. Об алгебраических методах в задачах распознавания и классификации . Распознавание, классификация, прогноз, 1988, т. 1, с. 9-16.

24. Журавлев Ю. И., Рязанов В. В., Сенько О. В. «Распознавание». Математические методы. Программная система. Практические применения. М.: Фазис, 2006.
25. Коршунов А., Белобородов И. и др. Определение демографических атрибутов пользователей микроблогов. Труды Института системного программирования РАН, 2013, том 25, с. 179-194.
26. Козлов В.Н. К математической теории зрительного восприятия. Интеллектуальные системы, 2006, том 10, в. 1-4, с. 141-166.
27. Кудрявцев В.Б., Андреев А.Е., Гассанов Э.Э. Теория тестового распознавания. – М.:ФИЗМАТЛИТ, 2007. – 320 с.
28. Куликова А.А. Подход к классификации пользователей в социальных сетях. Восточно-Европейский журнал передовых технологий, 2011, №2 (51), с. 246-452.
29. Смирнов В. «Большие данные» изменяют бизнес. CRN/RE, № 18 (413), 2013 г. Электронный ресурс (<http://www.crn.ru>)
30. Сойфер В.А. Компьютерная обработка изображений. Часть 2. Методы и алгоритмы // Соросовский образовательный журнал, 1996, №3, pp. 14-19.
31. Фролов А.Б. Классификация и распознавание топологических форм: учебное пособие для студентов. / Под ред. В. Б. Кудрявцева; М-во образования и науки Российской Федерации, Федеральное агентство по образованию, Московский энергетический ин-т (технический ун-т) 310-7/917, 2010.
32. Фролов А.Б. Принцип конечной топологии распознавания топологических форм // Известия РАН. Теория и системы управления, 2010, №1 – с. 68-76.
33. Фролов А.Б., Четрафилов И.Д. О некоторых подходах к распознаванию оптических образов текстов // Интеллектуальные системы, 1997, т.2, вып. 1-4, стр. 189-200.

34. Фу К. Структурные методы распознавания образов. М.: Мир, 1977. – 376 с.
35. Хуанг Т.С. и др. Быстрые алгоритмы в цифровой обработке изображений. М.: Радио и связь, 1984. – 224 с.
36. Abdi H., Valentin D., Edelman B., O'Toole A.J. More about the difference between men and women: Evidence from linear neural network and principal component approach // *Neural Computing*, 1995, vol. 7 (6), 1160–1164.
37. Alexandre L.A. Gender recognition: a multiscale decision fusion approach // *Pattern recognition letters*, 2010, vol. 31, pp. 1422-1427.
38. Anderson R., Stenger B., Wan V., Cipolla R. Expressive Visual Text-To-Speech Using Active Appearance Models. In: *IEEE International conference on computer vision and pattern recognition (CVPR'13)*, 2013, pp. 3382-3389.
39. Augusteijn M.F., Skujca T.L. Identification of human faces through texture-based feature recognition and neural network technology. In: *Proc. IEEE Conference on neural networks*, 1993, pp. 392-398.
40. Baluja S., Rowley H.A., Boosting sex identification performance // *International journal of computer vision*, 2007, vol. 71(1), pp. 111-119.
41. Basha A.F., Jahangeer G. Face Gender Image Classification Using Various Wavelet Transform and Support Vector Machine with various Kernels // *International Journal of Computer Science Issues (IJCSI)*, 2012, vol. 9(6), № 2, pp. 150-157.
42. Bellustin N., Kalafati Y. et al. Instant Human Face Attributes Recognition System // *International Journal of Advanced Computer Science and Applications: Special Issue on Artificial Intelligence*, 2011, pp. 112-120.
43. Ben S., Chen J., Su G. Piecewise linear aging function for facial age estimation. In: *16th IEEE International conference on image processing*, 2009, pp. 2753-2756.

44. Bekios-Calfa, J., Buenaposada, J.M., Baumela, L. Revisiting Linear Discriminant Techniques in Gender Recognition // *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, vol. 33(4), 858-864.
45. Bengio Y., Delalleau O., Simard C. Decision Trees do not Generalize to New Variations // *Computational Intelligence*, 2010, vol. 26(4), pp. 449-467.
46. Bengio Y., Glorot X. Understanding the difficulty of training deep feedforward neural networks. In: *Proc. AISTATS'10*, 2010, pp. 249-256.
47. Chang K., Chen C., Hung Y. A ranking approach for human age estimation based on face images. In: *Proc. 20th IEEE Conference on computer vision proceedings*, 2010, pp. 3396-3399.
48. Chang K., Chen C., Hung Y.P. Ordinal hyperplanes ranker with cost sensitivities for age estimation. In: *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, 2011, pp. 585-592.
49. Chen C., Ross A. Evaluation of gender classification methods on thermal and near-infrared face images In: *Proc. Of International joint conference on biometrics (IJCB)*, 2011, pp. 367-374.
50. Chen C., Gong S., Xiang T. Cumulative Attribute Space for Age and Crowd Density Estimation. In: *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR'13)*, 2013, pp. 2467-2474.
51. Cootes T., Edwards G., Taylor C. Active appearance models. *IEEE TPAMI*, vol. 23 (6) 2001, pp. 681-685.
52. Cootes T., Taylor C., Cooper D.H., Graham J. Active shape models – their training and application // *Computer vision and image understanding*, 1995, vol. 61 (10), pp. 38-59.
53. Cortes C., Vapnik V. Support-vector machine. *Machine Learning*, 1995, vol. 20, №. 3, pp. 273-297.

54. Demirkus M., Garg K., Guler S. Automated person categorization for video surveillance using soft biometrics. In: Proc. Biometrics technology for human identification (BTHI-VII), 2010, vol. 7667, pp. 564-576.
55. Deshmukh D. Gender recognition from model's face using SVM algorithm //International Journal of Engineering Trends and Technology (IJETT), 2014, vol. 10 № 1, pp. 13-18.
56. Edwards G., Taylor C., Cootes T. Face Recognition Using Active Appearance Models. In: Proceedings of the 5th European Conference on Computer Vision, 2006, vol. 2, pp. 581-595.
57. Etemad K., Chellappa R. Discriminant analysis for recognition of human face images // Journal of Optical Society of America., 1997, vol. 14, № 8, pp. 1724-1733.
58. FERET dataset: [Электронный ресурс]. Режим доступа: <http://www.nist.gov/itl/iad/ig/colorferet.cfm>
59. FG-NET Database. [Электронный ресурс]. Режим доступа: <http://www.cse.msu.edu/rgroups/biometrics/Publications/Databases/FGNETAgeEstimation.zip>.
60. Freeman W., Roth M. Orientation histograms for hand gesture recognition. Intl. Workshop on Automatic Faceand Gesture- Recognition. In: Proc. IEEE Computer Society, 1995, pp. 296–301.
61. Fu Y., Xu Y., Huang T.S. Estimating human ages by manifold analysis of face pictures and regression on aging features. In: Proc. IEEE Conference on multimedia and Expo, 2007, pp. 1383-1386
62. Gao W., Ai H. Face gender classification on consumer image in a multiethnic environment. In: International conference of biometrics (ICB), 2009, pp. 169-178.
63. Geng X., Zhou Z.H., Smith-Miles K. Automatic Age Estimation Based on Facial Aging Patterns // IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29(12), 2007, pp. 2234-2240.

64. Guo G., Fu Y., Huang T.S., Dyer C. A probabilistic fusion approach to human age prediction. In: Proc. IEEE CVPR- SLAM workshop, 2008.
65. Guo G., Fu Y., Huang T., Dyer C. Locally adjusted robust regression for human age estimation. IEEE Workshop on Applications of Computer Vision, 2008, pp. 376-382.
66. Guo G., Mu G., et al. A Study on Automatic Age Estimation using a Large Database. In: Proc. of IEEE International conference on computer vision (ICCV), 2009, pp. 1986-1991.
67. Golomb B.A., Lawrence D.T., Sejnovski T.J. «Sexnet»: A neural network identifies sex from human face. In: Proc. International conference on advances in neural information processing systems (NIPS-3), 1990, pp. 572-577.
68. Gutta S., Huang J., Jonathon P., Wechsler H. Mixture of Experts for Classification of Gender, Ethnic Origin, and Pose of Human Faces // IEEE Transactions on neural networks, 2000, vol. 11, № 4, pp. 948-960.
69. Felzenszwalb P.F., Girchik R.B., McAllester D., Ramanan D. Object detection with discriminatively trained part base models // IEEE Transactions on pattern analysis & machine intelligence, 2010, vol. 32(9), pp. 1627-1645.
70. Hadid A., Pietikainen M. Combining appearance and motion for face and gender recognition from videos // Pattern recognition letters, 2009, vol. 42 (11), pp. 2818–2827.
71. Harrington P. Machine Learning in Action. Manning Publication. 2012. – 307 p.
72. Heo J., Savvides M. Gender and Ethnicity Specific Generic Elastic Models from a Single 2D Image for Novel 2D Pose Face Synthesis and Recognition // IEEE Transactions on Pattern Analysis & Machine Intelligence, 2012, vol. 34, pp. 2341-2350.

73. Kalavi R. A study of cross-validation and bootstrapp for accuracy estimation and model selection // LJCAL. – 1995. – Pp. 1137-1145
74. Kazimov T., Mahmudova S. About a Method of Recognition of Race and Ethnicity of Individuals Based on Portrait Photographs // Intelligent Control and Automation, 2014, № 5, pp. 120-125.
75. Khan S.A., Nazir M., Riaz N. Gender Classification using Multi-Level Wavelets on Real World Face Images // Acta Polytechnica Hungarica, 2013, vol. 10, № 4, pp.193-208.
76. Khryashchev V., Priorov A., Shmaglit L., Golubev M. Gender Recognition via Face Area Analysis // Proceedings of the World Congress on Engineering and Computer Science (WCECS 2012), 2012, vol. 1.
77. Kilinc M., Akgul Y.S. Automatic human age estimation ising overlapped age group // Computer vision, imaging and computer graphics. Theory and application communications in computer and information science, 2013, vol. 359, pp. 313-325.
78. Labeled Faces in the Wild dataset. [Электронный ресурс]. Режим доступа: <http://vis-www.cs.umass.edu/lfw/>
79. Laptev I. Improvements of object detection using boosted histograms // Image and Vision Computing, 2009, vol. 27(5), pp. 535-544.
80. Lanitis A., Draganova C., Christodoulou C. Comparing different classifiers for automatic age estimation // IEEE Trans. On SMC-B, 2004, vol. 34(1), pp. 621-628
81. Latinis A., Taylor C., Cootes T. Automatic interpretation and coding of face images using flexible models // IEEE transactions on pattern analysis and machine intelligence, vol. 19(70), 1997, pp. 743-756.
82. Len B., Dat T., Xu H., Girija C. Face gender classification based on active appearance model and fuzzy k-nearest neighbors. In: The 2012 International Conference On Image Processing, Computer Vision & Pattern Recognition (IPVC), 2012, pp. 617-621.

83. Levi K., Weiss Y. Learning object detection from a small number of examples: the importance of good features. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2004, pp. 53–60.
84. Lian H., Lu B. Multi-view gender classification using local binary patterns and support vector machines. In: Proc. International symposium on neural networks, 2006, pp. 202–209.
85. Liu H., Lu J., Feng J., Zhou J. Group-aware Deep Feature Learning for Facial Age Estimation // Pattern Recognition, 2017, vol. 66, pp. 82–94.
86. Lu X., Chen H., Jain A.K. Multimodal facial gender and ethnicity identification. Advances in Biometrics. Lecture Notes in Computer Science, 2005, vol. 3832, pp. 554-561.
87. Luu K., Ricanek K., Bui T., Suen C. Age estimation using active appearance model and support vector machine regression. In: Proc. IEEE 3rd International conference on biometrics: theory, applications and systems (BTAS'09), 2009, pp. 1 – 5.
88. Maenpaa T. The Local binary pattern approach to texture analysis — extensions and applications. Oulu University Press, 2003.
89. Mahalingam G., Kambhamettu C. Face Verification with Aging Using AdaBoost and Local Binary Patterns. In: Proc. ICVGIP'10, 2010, pp. 46-55.
90. MALL dataset. [Электронный ресурс]. Режим доступа: http://www.eecs.qmul.ac.uk/~ccloy/files/datasets/mall_dataset.zip
91. Manning C.D., Raghavan P., Schütze H. Introduction to Information Retrieval. Cambridge, England: Cambridge University Press. – 2008.
92. Makinen E., Raisamo R. An experimental comparison of gender classification methods // Pattern recognition letters, 2008, vol. 29, pp. 1544-1556.

93. Moghaddam B., Yang M., Learning gender with support faces // IEEE Transactions and pattern analysis and machine intelligence, 2002, vol 24 (5), pp. 707-711.
94. MORPH (Craniofacial Longitudinal Morphological Face Database), [Электронный ресурс]. Режим досту <http://www.faceaginggroup.com/morph/>
95. Nielsen M.A. Neural networks and deep learning. Determination Press, 2015.
96. Ojala T., Pietikainen M., Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns // IEEE Transactions on pattern analysis and machine intelligence. vol. 24 (7), 2002, pp. 971-987.
97. Pannachiotti M., Popescu A. A machine learning approach to Twitter user classification. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 2011, pp.281-288.
98. Philips P.J., Moon H., Rauss P., Rizvi S.A. The FERET September 1996 database and evaluation procedure. In: First international conference on audio and video-based biometric person authentication, 1997, pp. 1- 8.
99. Qiu X., Sun Zh., Tan T. Global Texture Analysis of Iris Images for Ethnic Classification // Lecture Notes in Computer Science, 2005, vol. 3832, pp. 411-418
100. Rahman M.H., Das T., Sarnaker M. Face Detection and Sex Identification from Color Images using AdaBoost with SVM based Component Classifier // International Journal of Computer Applications, 2013, vol. 76, № 3, pp. 1-6.
101. Ramanathan N., Chellappa R. Modeling age progression in yang faces. In: Proc. IEEE Conference on computer vision and pattern recognition (CVPR'06), 2006, pp. 387-394.

102. Ren H., Li Z.N. Age estimation based on complexity-aware features. In: Proc. Asian conference on computer vision, 2014, pp. 1-14.
103. Saatci Y., Town C. Cascaded Classification of Gender and Facial Expression using Active Appearance Models. In: Proc. 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2006, pp. 393-400.
104. Shakhnarovich G., Viola P.A., Moghaddam B., A unified learning framework for real time face detection and classification. In: Proc. IEEE International conference on automatic face & gesture recognition, 2002, pp. 14-21.
105. Shan C. Learning local binary patterns for gender classification on real-world face images // Pattern Recognition Letters, 2012, vol. 33(4), pp. 431-437.
106. Tamura, S., Kawai, H., Mitsumoto, H. Male/female identification from 8 to 6 very low resolution face images by neural network // Pattern Recognition, 1996, vol. 29 (2), pp. 331–335.
107. Tariq U.U., Ahmad W., Asif M.D., Hassan M. Gender Perception From Faces Using Boosted LBPH (Local Binary Patten Histograms) // Carpathian Journal of Electronic and Computer Engineering, 2013, vol. 6(1), pp. 8-15.
108. Van Rijsbergen C. Information Retrieval, 2nd ed. – Butterworths, London, UK. – 1979.
109. Vapnik V. Statistical learning theory. John Wiley, New York, 1998.
110. Yilionias J., Hadid A., Hong X., Pietikainen M. Age estimation using locale binary patterns kernel density estimate. In: Proc. International conference on image analysis and processing (ICIAP'13), 2013, pp. 141-150.
111. Yang Z., Ai H. Demographic classification with local binary patterns. In: Proc. IEEE International conference of biometrics, 2007, pp. 464-473.

112. Yang Z., Li M., Ai H. An experimental study on automatic face gender classification. In: Proc. International Conference on Pattern Recognition (ICPR), 2006, pp. 1099–1102.
113. Zamvar S.C., Ladhake S.A., Ghate U.S. Human Face Detection and Tracking for Age Rank, Weight and Gender Estimation based on Face Images utilizing Raspberry Pi Processor // International Research Journal of Engineering and Technology (IRJET), 2017, v. 4 (2), pp. 185-188.
114. Adaboost Python Library [Электронный ресурс]. Режим доступа: <https://Pythonhosted.org/milk/adaboost.html>
115. SVM Python Library [Электронный ресурс]. Режим доступа: <https://www.csie.ntu.edu.tw/~cjlin/libsvm>
116. Butcher C., Clarke G., Cormack G. Information Retrieval: Implementing and Evaluating Search Engines. MIT Press. 2010.
117. Manning C., Raghavan P., Schütze H. Introduction to information retrieval. Cambridge University Press. 2008.
118. Stegmann M.B. Analysis and Segmentation of Face Images using Point Annotations and Linear Subspace Techniques. Informatics and Mathematical Modelling, Technical University of Denmark Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, Denmark IMM Technical Report IMM-REP-2002-22, August 2002, pp. 25. URL: http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/922/pdf/imm922.pdf
119. Рыбинцев А.В., Лукина Т.М., Конушин В.С., Конушин А.С. Модификация алгоритма классификации возраста людей по изображению лица на основе метода ранжирования // Графикон-2012: 22-ая международная конференция по компьютерной графике: Тез. докл. – Москва, 2012. – С. 133-137.
120. Рыбинцев А.В., Лукина Т.М., Конушин В.С., Конушин А.С. Возрастная классификация людей по изображению лица на основе

- метода ранжирования и локальных бинарных шаблонов // Системы и средства информатики. – 2013. – Том 23, № 2, – С. 48-59.
121. Рыбинцев А.В., Конушин В.С., Конушин А.С. Последовательная половая и возрастная классификация людей по изображению лица на основе ранжированных локальных бинарных шаблонов // Компьютерная оптика. – 2015. – Том 39, № 5, - С. 762-769.
122. Рыбинцев А.В. Исследование подходов к определению атрибутов личности по изображению лица методами машинного зрения. Национальный исследовательский университет «МЭИ». М., 2015. – 57 с. Деп. в ВИНТИ, № 86 – В2016.
123. Рыбинцев А.В. Снижение размерности пространства LBP-признаков в задаче определения атрибутов личности по изображению лица // Вестник МЭИ. – 2016. - № 1, - С. 33 - 38.
124. Рыбинцев А.В. Использование процедуры усеченного бутстраппинга в задаче определения атрибутов личности по изображению лица // 22-ая международная научно-техническая конференция «Радиоэлектроника, электротехника, энергетика»: Тез. докл. – Москва, 2016. Том 1, с. 210-211.
125. Рыбинцев А.В. Последовательное определение атрибутов личности по изображению лица на основе ранжированных LBP-признаков // Вестник МЭИ. – 2017. - № 5, - С.121-129. DOI: 10.24160/1993-6982-2017-5-121-129 (Из списка ВАК)
126. Rybintsev A. Age Estimation from a Face Image in a Selected Gender-Race Group based on Ranked Local Binary Patterns // Complex & Intelligent Systems, 2017, Volume 3, Issue 2, pp. 93-104. DOI: 10.1007/s40747-017-0035-y (Из списка ВАК, решение Президиума ВАК № 9\11)

Приложение

Акт о внедрении результатов диссертационной работы

«УТВЕРЖДАЮ»

Технический директор

ООО «ДаблЭй»

Беляев А.Ю.

2018 г.



А К Т

**о внедрении результатов диссертации Рыбинцева А.В.
на тему «Исследование, модификация и разработка методов
компьютерного зрения для задач определения атрибутов личности
по изображению лица»**

Настоящим удостоверяется, что результаты диссертации Рыбинцева А.В. на тему «Исследование, модификация и разработка методов компьютерного зрения для задач определения атрибутов личности по изображению лица» обладают актуальностью, представляют практический интерес и были внедрены в компании ООО «ДаблЭй» при реализации проекта «Разработка программного обеспечения для автоматической возрастной классификации информационной продукции», выполняемого при организационной и финансовой поддержке Федерального государственного бюджетного учреждения «Фонд содействия развитию малых форм предприятий в научно-технической сфере» (программа «Старт», проект С1-19968, направление Н1: Информационные технологии).

Использование предложенных Рыбинцевым А.В. алгоритмов описания и классификации изображений позволили в разрабатываемом программном обеспечении:

- повысить быстродействие, качество описания и классификации размещенной в сети Internet графической информации;
- включить в состав программного комплекса модуль, отвечающий за допуск людей к информации, имеющей возрастные ограничения.

Начальник отдела
разработки ПО



Симбирцева М.С.