



ТЕХНИЧЕСКИ УНИВЕРСИТЕТ – СОФИЯ
ФАКУЛТЕТ ПО ТЕЛЕКОМУНИКАЦИИ
КАТЕДРА "РАДОИКОМУНИКАЦИИ И
ВИДЕОТЕХНОЛОГИИ"

**ПРИЛОЖЕНИЕ НА СТАТИСТИЧЕСКИТЕ
ЛИНЕЙНИ ПРЕОБРАЗУВАНИЯ,
РЕАЛИЗИРАНИ С ЯДРО ПРИ
РАЗПОЗНАВАНЕТО НА ИЗОБРАЖЕНИЯ
НА ОБЕКТИ**

ДИСЕРТАЦИОНЕН ТРУД ЗА ПРИДОБИВАНЕ НА
ОБРАЗОВАТЕЛНО-НАУЧНА СТЕПЕН "ДОКТОР"

Автор:

МАГ. ИНЖ.
КРАСИМИР ТОНКОВ
ТОНЧЕВ

Ръководител:

ПРОФ. Д-Р ИНЖ.
ОГНЯН ЛЮБЕНОВ
БУМБАРОВ

Професионално направление: 5.3. КОМУНИКАЦИОННА И
КОМПЮТЪРНА ТЕХНИКА

Научна специалност: ТЕЛЕВИЗИОННА И
ВИДЕОТЕХНИКА

София, 2022

Благодарности

Издавам искрената си благодарност към моя ръководител, проф. д-р инж. Огнян Бумбаров за оказаното съдействие и насоки в изготвянето на този дисертационен труд. Благодаря, също така, и на проф. д.т.н. Владимир Пулков и доц. д-р Агата Манолова за ценните съвети, помощ и кооперация в проведените научни изследвания и оформянето на дисертацията.

Благодаря на моите колеги от университета за успешната ни съвместна работа, разбирателство, съвети и хубави моменти.

Дълбоко съм признателен на майка ми, за неотлъчната ѝ подкрепа и насърчение през тези години.

Отправлям и искрена благодарност на близките ми приятели, които ме поддържаха по време на работата ми.

Списък на фигурите

1.1	Обща блок-схема за анализ на изображения на обекти.	4
1.2	Условно разделяне на признаковите пространства в компютърното зрение: а) в зависимост от използване на данните; б) в зависимост от приложението на оператора върху областта на изображението; г) в зависимост от типа на оператора; д) в зависимост от типа на изходното пространство.	9
1.3	Визуализация на изображението Φ от пространството на лицата в хилбертово пространство образувано от гаусови функции	24
2.1	Блок-схема на предложеният алгоритъм за класификация на изображения на лица във възрастни групи - фаза Обучение.	38
2.2	Блок-схема на алгоритъм за класификация на изображения на лица във възрастни групи - фаза Тест.	40
2.3	Архитектура на конволюционната невронна мрежа LeNet-5 [110].	43
2.4	Визуализация на процеса по подравняване на лица.	44
2.5	Примерна диаграма на пълносвързаният граф определящ РСА съставен от 6 възела.	46
2.6	Класификация с хиперравнина в два класа: А) разделящи хиперравнини; Б) оптимална хиперравнина с максимално разстояние.	51
3.1	Блок-схема на алгоритъм за разпознаване на изражения на лица с избор на log-Gabor признаци - фаза Обучение.	60
3.2	Блок-схема на алгоритъм за разпознаване на изражения на лица с избор на log-Gabor признаци - фаза Тест.	63
3.3	log-Gabor филтър в честотна област.	64
3.4	Блок-схема на алгоритъм за разпознаване на изражения на лица с геометрични признаци и метода с ядрото, реализиран за големи обеми от данни - фаза Обучение.	69
3.5	Блок-схема на алгоритъм за разпознаване на изражения на лица с геометрични признаци и метода с ядрото, реализиран за големи обеми от данни - фаза Тест.	71
3.6	Илюстрация, поясняваща целта на метод за детектиране на ключови точки по лицеви признаци.	72
4.1	Блок-схема на алгоритъм за разпознаване на 3D изображения на лица, представени с дълбочинно изображение - фаза Обучение.	85
4.2	Визуализация на точков шум на 3D изображение на лице.	86
4.3	Сканирано 3D изображение и извлеченото дълбочинно изображение.	86
4.4	Блок-схема на алгоритъм за разпознаване на 3D изображения на лица описваща фазата Тест.	88

4.5	Блок-схема на алгоритъм за класификация на 3D изображения на лица с множество ядра - фаза Обучение.	93
4.6	Блок-схема на алгоритъм за класификация на 3D изображения на лица с множество ядра - фаза Тест.	96
5.1	Примерни изображения от базата данни FG-NET.	104
5.2	Примерен базис на PCA и SR базис, генериран с изображения на лица.	106
5.3	Примерни изображения от базата данни СК+.	108
5.4	Разпознаването на израженията като функция от броя на признаците избрани от алгоритъма за избор на признаци.	109
5.5	Примерни изображения от базата данни SFEW 2.0.	111
5.6	Примерна извадка от 3D сканирани лица от базата данни SHREC.	113
5.7	Примерна извадка от 3D сканирани лица от базата данни SHREC.	115
5.8	Оригинално 3D сканирано изображение от SHREC, и лицевият регион след предварителната обработка.	116
5.9	Примерна лицева повърхнина със съответните ѝ нормални вектори към нея.	116
5.10	Визуализация на признаците LARK.	117

Списък на таблиците

1.1	Характеризация на признаковите пространства в зависимост от операторите които ги генерират.	14
1.2	Характеризация на признаковите пространства в зависимост от операторите реализирани с дълбоки невронни мрежи.	15
5.1	Параметри на алгоритъма за определяне на възраст в групи.	105
5.2	Коефициент на разпознаване (ефективност) на предложения алгоритъм за различните конфигурации.	106
5.3	Брой изображения за различните изражения в СК+.	107
5.4	Параметри на алгоритъма за разпознаване на изражения с избиране на log-Gabor признаци.	108
5.5	Резултати от експериментите за определяне на ефективност на предложения алгоритъм и сравнението му с други.	109
5.6	Параметри на алгоритъма за разпознаване на изражения, използвайки апроксимация на метода с ядрото за големи обеми от данни.	111
5.7	Резултати от експериментите за определяне на ефективност на предложения алгоритъм и сравнението му с други.	112
5.8	Параметри на алгоритъма за разпознаване на 3D лица с дълбочинно изображение.	114
5.9	Резултати от експериментите на предложения алгоритъм, реализиран с различни класификатори и различни обучаващи извадки.	114
5.10	Параметри на алгоритъма за разпознаване на 3D лица с множество ядра.	116
5.11	Резултати от експериментите за определяне на ефективност на предложения алгоритъм и сравнението му с други.	117

Списък на съкращенията

- 6G** Sixth Generation Mobile Networks
- AAM** Active Appearance Models
- AGES** AGing pattErn Subspace
- ARS** Angular Radial Signature
- ASM** Active Shape Models
- BIF** Bio Inspired Features
- BoW** Dense Bag Of Words
- CEA** Conformal Embedding Analysis
- CLM** Constrained Local Models
- CNN** Convolutional Neural Network
- DoG** Difference of Gaussians
- GDA** Generalized Discriminant Analysis
- GH** Gromov-Hausdorf
- GMM** Gaussian Mixuture Models
- HCI** Human Computer Interface
- HOG** Hisogram of Oriented Gradients
- ICA** Independent Component Analysis
- ICP** Iterative Closest Point
- KDDA** Kernel Direct Discriminant Analysis
- KKT** Karush-Kuhn-Tacker
- KL** Kullback-Leibler
- KLDA** Kernel LDA
- KLPP** Kernel LPP
- k-NN** k-Nearest Neighbors

KPCA Kernel PCA

LAR Least Angle Regression

LARK Locally Adaptive Regression Kernels

LBP Local Binary Patterns

LDA Linear Discriminant Analysis

LDE Locally Discriminant Embedding

LE Laplacian Eigenmaps

LLE Locally Linear Embedding

LOPO leave-one-person-out

LPP Locality Preserving Projections

LPQ Local Phase Quantisation

LTP Local Ternary Patterns

MAC Multiply and accumulate

MCFS Multi-Cluster Feature Selection

MDS Multi Dimensional Scaling

MFA Marginal Fisher Analysis

NMF Non-negative Matrix Factorization

NPP Neighborhood Preserving Projections

NTK Neural Tangent Kernel

OLPP Orthogonal LPP

PB SIFT Part Based SIFT

PCA Principal Component Analysis

PCL Point Cloud

QLZM Quantized Local Zernike Moments

ReLU Rectified Linear Unit

RMSE Root Mean Square Error

SAG Stochastic Averaged Gradient

SDM Supervised Descent Method

SFP Spatially Flexible Patch

SGD Stochastic Gradient Descent
SIFT Scale Invariant Features Transform
SR Spectral Regression
SSD Self-Similarity Descriptor
SURF Speed-Up Robust Features
SVD Singular Value Decomposition
SVM Support Vector Machines
SVR Support Vector Regression
ULTP Uniform Local Ternary Patterns
WLD Weber Local Descriptors
ROC Receiver Operating Characteristics
SFEW Static Facial Expressions in the Wild

Списък на означенията

\mathbb{R}	множеството на реалните числа
\mathbb{N}	множеството на естествените числа
\mathbb{R}^n	евклидово пространство
\mathcal{M}	многообразие
\mathcal{H}	хилбертово пространство
γ	вектор съдържащ класовете в задачата за разпознаване
γ	клас принадлежащ на γ , естествено число
$\hat{\gamma}$	определен клас от класификатора
\mathcal{X}	множеството на 2D и/или 3D изображенията
X	изображение
\mathcal{F}	множество от признаци
Φ	оператор с който се извличат признаци
$\{\Phi, \mathcal{F}\}$	признаково пространство
α, β	скалари с общо предназначение
d	размерността на признаковото пространство което се разглежда
\mathbf{X}	матрица съдържаща изображения по колони
$X_{i,j}$	елемент от матрицата \mathbf{X}
\mathbf{W}	тегловна матрица
$W_{i,j}$	елемент от матрицата \mathbf{W}
L	матрицата graph-Laplacian
G	граф
$k(\cdot, \cdot)$	ядро
$\phi(\cdot)$	изображение от пространството на лицата в хилбертово пространство
δ_x	функционал на Дирак в точката x
$\langle \cdot, \cdot \rangle_{\mathcal{H}}$	скалярно произведение за хилбертовото пространство \mathcal{H}
\mathcal{B}_{rgb}	множество от цветни RGB 2D изображения
\mathbf{x}_μ	вектор представляващ средна стойност
$\bar{\mathbf{X}}$	матрица центрирана по колони
\mathbf{C}	ковариационна матрица
\mathbf{E}	идентитет, матрица (единичната матрица)
\mathbf{e}	единичен вектор
$\bar{\mathbf{V}}$	собствени вектори на ковариационната матрица
Λ	матрица съдържаща собствени стойности по диагонала
\mathbf{V}	собствени вектори на ковариационната матрица
\mathbf{F}	матрица съдържаща признаци по колони
\mathbf{f}	признаков вектор
\mathbf{A}	матрица с решенията на регресията на алгоритъма Spectral Regression
\mathbf{K}	матрица на Грам
$K_{i,j}$	елемент от матрицата на Грам
α	опорни вектори

\mathcal{B}_{3D}	множество от 3D изображения на лица
\mathcal{D}	дълбочинно изображение
\mathbb{E}	оператора математическо очакване
$p_G(\cdot)$	гаусово вероятностно разпределение
$p_L(\cdot)$	лапласово вероятностно разпределение
\mathbf{S}_B	матрица на вътрешно-класовите вариации
\mathbf{S}_W	матрица на между-класовите вариации
k_R	резултантно ядро което е изпъкнала комбинация от множество ядра

Съдържание

Благодарности	i
Списък на фигурите	i
Списък на таблиците	iii
Списък на съкращенията	iv
Увод	xiii
Структура на дисертацията	xiv
1 Анализ на състоянието на проблема по литературни данни	1
1.1 Анализ на изображения на обекти в компютърното зрение	1
1.1.1 Визуална информация от човешки лица	2
1.1.2 Общи положения на разпознаването на изображения на обекти	3
1.2 Признакови пространства използвани при анализ на изображения на обекти и лица	5
1.2.1 Класификация на признаковите пространства	8
1.2.2 Признакови пространства от тип линейни под-пространства и нелинейни пространства	13
1.2.3 Намаляване на размерността на признаковите пространства	16
1.2.4 Признакови пространства генерирани от линейни статистически преобразувания	16
1.2.5 Признакови пространства от тип многообразие	19
1.2.6 Обща рамка за моделиране на линейни статистически преобразувания	22
1.3 Методи в разпознаването на изображения на обекти, реализирани с ядро	23
1.3.1 Имплицитна реализация на нелинейност чрез методът с ядрото	24
1.3.2 Линейни статистически трансформации, реализирани с ядро, използвани в разпознаването на изображения на обекти	26
1.4 Изводи	27
1.5 Дефиниране на целта и основните задачи на настоящата дисертация	28
2 Разработване на алгоритъм за определяне на възраст от изображения на лица	30
2.1 Анализ на литературните данни по определяне на възраст от изображения на лица	31
2.1.1 Изводи от литературния анализ по определяне на възраст от изображения на лица	35

2.2	Класификация по възрастови групи с линейни статистически преобразувания, реализирани по метода с ядрото	36
2.2.1	Блок-схема на потока на информацията и анализ на предложения алгоритъм	37
2.2.2	Детектиране на лица в изображения на естествени сцени с комбиниран класификатор	41
2.2.3	Регистрация на изображения на лица	44
2.2.4	Намаляване на размерността с Principal Component Analysis	45
2.2.5	Извличане на холистични признаци с алгоритъма Spectral Regression	47
2.2.6	Класифициране във възрастови групи с Support Vector Machines	50
2.3	Основни резултати и изводи	53
2.4	Приноси към Глава 2	54
3	Разработване на метод и алгоритми за разпознаване на изражения на лица	55
3.1	Анализ на литературните данни по разпознаване на изражения на лица	56
3.1.1	Изводи от литературния анализ по разпознаване на изражения на лица	58
3.2	Разпознаване на изражения на лица чрез избор на признаци в log-Gabor признаково пространство	59
3.2.1	Блок-схема на потока на информацията и анализ на предложения алгоритъм	59
3.2.2	Извличане на признаци с log-Gabor филтри	62
3.2.3	Избор на признаци чрез разрежена регресия	65
3.3	Разпознаване на изражения на лица за големи обеми от данни реализирани по метода с ядрото	67
3.3.1	Блок-схема на потока на информацията и анализ на предложения алгоритъм	68
3.3.2	Извличане на геометрични признаци с Supervised Descent Method	72
3.3.3	Разпознаване на изражения с метода с ядрото реализирани за големи обеми от данни	74
3.4	Основни резултати и изводи	77
3.5	Приноси към Глава 3	78
4	Разработване на метод и алгоритми за разпознаване на 3D изображения на лица	80
4.1	Анализ на литературните данни по разпознаване на 3D изображения на лица	81
4.1.1	Изводи от литературния анализ по разпознаване на 3D изображения на лица	82
4.2	Разпознаване на 3D лица чрез дълбочинно изображение и линейни статистически трансформации	83
4.2.1	Блок-схема на потока на информацията и анализ на предложения алгоритъм	84
4.2.2	Формиране на дълбочинно изображение	89

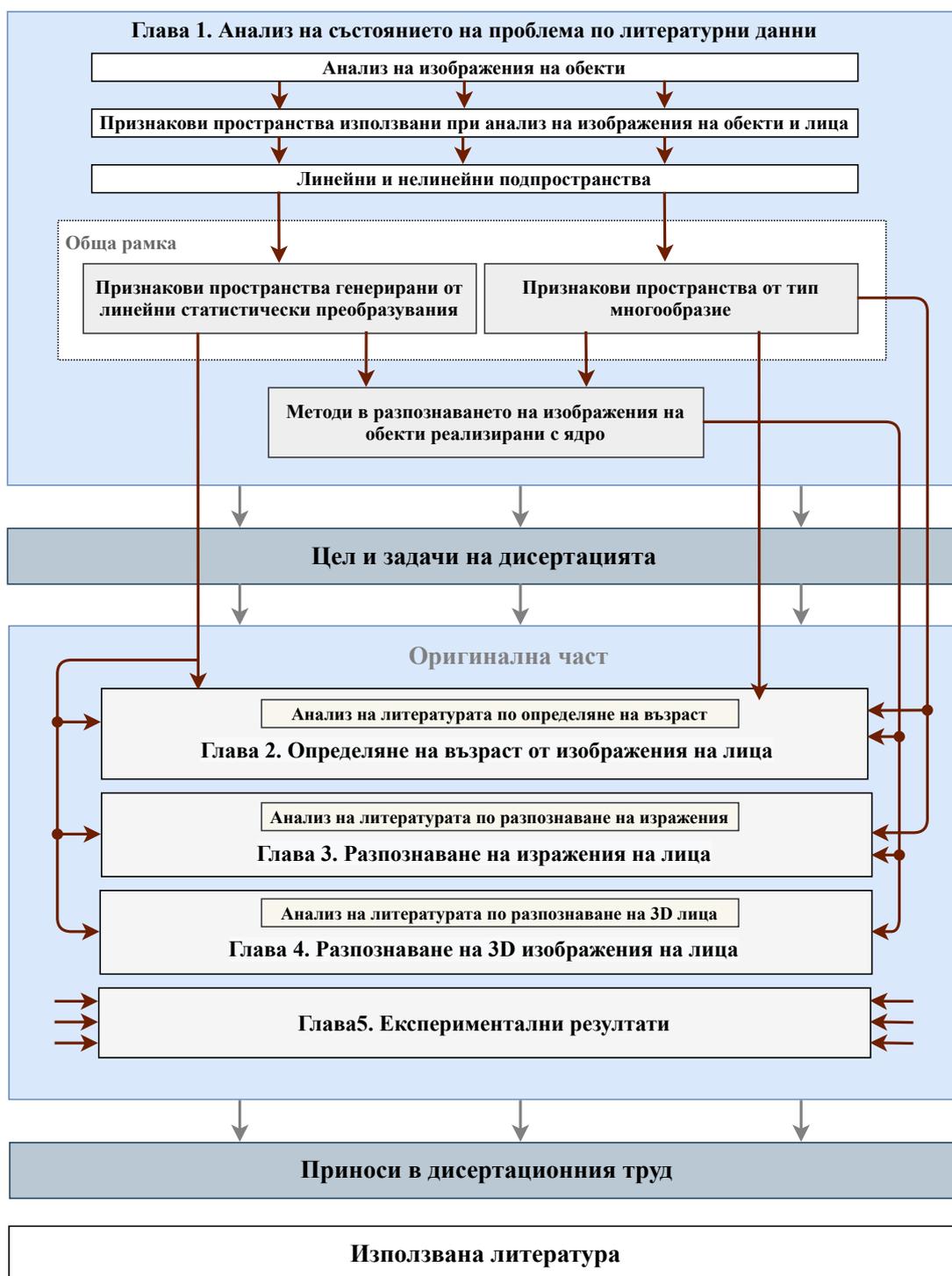
4.2.3	Извличане на признаци от дълбочинно изображение чрез кривината на формата	89
4.2.4	Извличане на дискриминативни признаци с Linear Discriminant Analysis	90
4.3	Разпознаване на 3D лица с комбиниране на допълващи се локални признаци чрез множество ядра	91
4.3.1	Блок-схема на потока на информацията и анализ на предложения алгоритъм	92
4.3.2	Извличане на признаци чрез нормали към повърхнина	95
4.3.3	Извличане на признаци чрез Locally Adaptive Regression Kernels	97
4.3.4	Класификация на 3D лица чрез множество ядра	98
4.4	Основни резултати и изводи	100
4.5	Приноси към Глава 4	101
5	Експериментални резултати от разработените алгоритми	103
5.1	Експериментални изследвания на алгоритмите към Глава 2	103
5.1.1	Експериментални резултати от теста на алгоритъма за определяне на възраст в групи	103
5.2	Експериментални изследвания на алгоритмите към Глава 3	107
5.2.1	Експериментални резултати от алгоритъма за разпознаване на изражения с избор на признаци в log-Gabor признаково пространство	107
5.2.2	Експериментални резултати от алгоритъма за разпознаване на изражения на лица за големи обеми от данни, реализиран по метода с ядрото	110
5.3	Експериментални изследвания на алгоритмите към Глава 4	112
5.3.1	Експериментални резултати от алгоритъма за разпознаване на 3D лица чрез представяне с дълбочинно изображение	112
5.3.2	Експериментални резултати от алгоритъма за разпознаване на 3D лица с комбиниране на допълващи се локални признаци чрез множество ядра	115
5.4	Основни резултати и изводи	117
5.5	Приноси към Глава 5	118
	Приноси в дисертационния труд	120
	Авторски публикации по дисертационните изследвания	122
	Участие в научно-изследователски проекти по време на докторантурата	123
	Списък на цитиранията	124
	Заклучение и бъдеща работа	125
	Използвана литература	126

Увод

Напредъкът на компютърните технологии през последните години доведе до все повече приложения, при които компютърът и автоматизираните системи вземат решения вместо човек. Автономни автомобили, анализ на медицински изображения, биотехнологиите и роботиката са само част от всички приложения при които човек бива заместван от компютър, за да се постигне по-голяма точност, по-висока производителност и съответно, по-добро качество. Несъмнено в основата на тези приложения стои разпознаването на обекти от изображения и видеопоследователности, които през последните години, благодарение на невронните мрежи, доведоха до бум в научно-изследователската работа и използването им в индустрията. В основата на този бум стои и ускоряването на изчисленията благодарение на все по-достъпните и все по-бързи специализирани изчислителни машини. Нещо повече, с появата на невро-морфните чипове се очаква да се достигнат плътности на изчислителните елементи, сходни с тези на човешкия мозък, което ще стои в основата на нови и по-добри алгоритми за разпознаване на обекти и появата на подобрени автоматизирани решения, които са били немислими само допреди две десетилетия.

Настоящата дисертация разглежда анализа на изображения на човешки лица, който се явява основен елемент в комуникацията между човек и машина, като са представени резултати свързани с три приложения, а именно, определяне на възраст, разпознаване на изражения на лица и разпознаване на лица. Автоматичното определяне на възраст от компютъра би подпомогнало машината да вземе решение спрямо възрастта на човек затова, например, дали човек е в определена възрастова граница да получи достъп до определена информация. Разпознаването на емоции и изражения би подпомогнало по-точно да се определи настроението което е особено полезно в медицински приложения. Разпознаването на лица намира все по-голямо приложение в системите за достъп, които изискват по-голяма точност, по-малка възможност за заблуда на машината и по-бързо вземане на решение.

Структура на дисертацията



1 Анализ на състоянието на проблема по литературни данни

1.1 Анализ на изображения на обекти в компютърното зрение

Компютърното зрение е област от компютърните науки целяща да възпроизведе човешката визуална система чрез сензори, компютри и специализирани алгоритми. Задачите които се решават са разнообразни и често границите между тях са размити, като една задача може да се причисли към два различни класа от задачи. Основните класове които могат да се отделят са следните: реконструиране на 3D модели на обекти; сегментация на сцени; генериране на изображения и разпознаване на изображения на обекти. Последният клас от задачи е свързан с текущата дисертация, като в него се разглеждат проблеми свързани с класификация и регресия при дадено входно изображение. При класификацията се цели даден обект автоматично да се причисли към предварително зададени класове, докато при регресията се цели да се определи някаква количествена оценка за обекта или да се извлече семантична информация. С напредването на изчислителната техника и изследователската дейност, съвременните алгоритми за компютърно зрение постигат резултати (в определен клас задачи) съизмерими или дори по-добри от човешката визуална система [1]. Въпреки това, голямото разнообразие от реални ситуации, в които трябва да работят алгоритмите оставя голям брой нерешени задачи. Разпознаването на изображения на обекти остава отворен проблем с много приложения. Детектирането и разпознаването на пътни знаци, пътна маркировка, автомобили в съседство и динамиката на сцената, са само част от основните компоненти на съвременните автономни автомобили които използват алгоритми за разпознаване на обекти. Голям клас от задачи в медицината също се решават чрез разпознаването на обекти. Определянето на тип заболяване от ядрено-магнитен резонанс изображения или белодробно заболяване от рентгенова снимка са направления в медицината, които все по-често се решават от компютърни системи използващи разпознаване на обекти. Не на последно място, приложение намират и в системи за взаимодействие между човек и компютър. Разпознаването на емоциите и емоционалното състояние, поведението и идентичността на човек са част от алгоритмите, от които компютърната система се нуждае за да може да взаимодейства по-добре с човек.

1.1.1 Визуална информация от човешки лица

Разпознаването на лица е задача, която хората извършват ежедневно, рутинно и без усилие. При бебетата, един от първите визуални стимули е свързан с човешките лица. От най-ранна възраст, човек се учи да идентифицира, определя, и в крайна сметка да различава тези хора, които са близки и важни за него, чрез техните лица. Това умение остава с човека през целия му живот, поради което, той го възприема като даденост. Въпреки това лицето освен идентичността, само по себе си, носи и друга полезна информация. Например, речта е придружена от лицево движение, в резултат на което, четенето по устните е възможно. Също така, израженията на лицето се дължат на видима деформация на мускулите му, които дават възможност да се оцени настроението, характера и умисъла на събеседника. Човек постоянно „чете“ по лицата на другите и този процес играе важна роля в социалното взаимодействие и комуникация. Моментните изменения във външния вид на лицето се случват в доста кратък времеви период, от порядъка на секунди или милисекунди, и носят информация свързана с речта и настроението. По-трайните визуални изменения на лицето позволяват да се оценят фактори като възраст, пол, етнически произход и идентичност.

Изразът на лицето, регистриран от околните е видима проява на: емоционално състояние, когнитивна активност, намерение, личност и психопатология на човек [2]. То играе основна комуникативна роля в междуличностните отношения. Изразът на лицето предава невербално комуникационни знаци при взаимодействия между хората. Те могат също да допълват речта като помагат на слушателя да извлече и засили значението на произнесените думи. Както е посочено в [3], израженията на лицето имат значителен ефект върху слушащия събеседник, като лицето на говорещия предава около 55% от емоцията, други 38% от нея се предават чрез гласова интонация, а останалите 7% от изговорените думи.

Човешките лица, сами по себе си, представляват уникален клас обекти. Те имат много сходна структура и лицевите им характеристики винаги са конфигурирани по доста сходен начин. Въпреки това, предвид тяхната прилика е забележително, че човек може да разграничи толкова много различни хора въз основа, само на външния вид на лицето. Още по-забележително е, когато се вземе в предвид, че разпознаването на хората по техните лица е силно динамично, както в пространството, така и във времето, и силно зависи от даденият контекст. Съдейки по човешкото възприятие и поведение е установено, че човек използва предимно статичните характеристики в разграничаването между различните лица.

От научна гледна точка трябва да се подчертае, че възприемането на лицата е пространно-времева задача. Човек може да идентифицира някого от статична снимка, но в реалния свят обикновено той наблюдава и идентифицира хората по техните лица в динамична среда. Обликът на лицето се променя поради относителното движение на лицето и наблюдателя. Промените в осветлението и околната среда също оказват влияние върху възприетата информация от лицето. Ето защо, възприятието на лицето е нещо повече от възприемането на статични снимки на лица. По-точно, човек има способността да открива и следи лица, докато се движат през сложни и динамични сцени. Поради това, процесът на идентифициране на човек по лицето му в присъствието на противоречиви фактори, които променят външния вид на лицето, е трудна задача. Ето защо е важно да се проучат и анализират основните причини за промяна на външния вид на лицето.

В повечето случаи, лицето може да промени външния си вид поради присъщи или външни фактори. Промените дължащи се на присъщи фактори са независими от наблюдателя и се дължат единствено на физическото естество на лицето. Промени от външни фактори от друга страна, възникват когато лицето е наблюдавано в различни условия на осветеност или от различен ъгъл. В общи линии, лицата имат много степени на присъщи промени, които са трудни за характеризирание аналитично - като например лицевото изражение, пол, възраст и идентичност. Едва когато изследователите се опитват да дублират човешките умения за анализиране на лицето чрез компютърна система, като например да бъдат идентифицирани лица или да се разпознаят техните изражения, възраст и пол, може да се осъзнае сложността на тази задача. Това е довело до въвеждането на множество различни подходи за решаването ѝ, и въпреки, че е постигнат голям напредък, много проблеми все още остават неразрешени. Широкото разпространение на мощни евтини настолни и вградени изчислителни системи през последните години, генерира огромен интерес в научните среди към анализа на човешки лица от изображения. Това от своя страна доведе до създаването на редица софтуерни приложения за автоматичното разпознаване на хора по техните лица или на тяхното емоционално състояние, както и определянето на пол, възраст и етническа принадлежност [4]. Алгоритмите анализират цифрови изображения и видеоследователности на човешки лица за целите на биометричната автентификация, наблюдение, взаимодействие между човек и машина, и мултимедийно управление.

Интереса към анализа на човешките лица, и в частност, на техните характеристики също така се дължи и на факта, че те имат няколко предимства в сравнение с други биометрични технологии: лицето е естествено, ненаатрапчиво, и достъпно чрез съвременните методи за заснемане. Лицето, като триизмерен обект, подложено на променящо се осветление, поза, изражение и т.н., трябва да бъде идентифицирано въз основа на неговото двуизмерно или триизмерно изображение. Информацията, която носят лицето и неговите изражения може да играе важна роля навсякъде, където хората взаимодействат с машини. Автоматично разпознаване на лицеви изрази може да действа като компонент на интерфейси човешка-машина [5] или с други думи, за да помогне на компютъра да поеме социална роля като инструктор или помощник [6].

1.1.2 Общи положения на разпознаването на изображения на обекти

Анализирането на изображенията на обекти от компютрите преминава през много етапи на обработка преди да се извлече необходимата информация. Нуждата от такава комплексна обработка се дължи на големият брой вариации, характерни при заснемането на реални сцени. По-точно, в засенените изображения освен обекта на интерес, се съдържат и други обекти, които са част от сцената и трябва да бъдат премахнати. Също така, вариации свързани с осветеността оказват голямо влияние и те трябва да бъдат нормализирани към този тип вариации. Друг надежден подход е те да бъдат взети предвид при извличането на признаци, което се среща при използването на съвременни невронни мрежи. Вариациите на позата на обекта представляват друг фактор който също оказва съществено влияние при анализа, защото важни характеристики на обекта могат да бъдат

пропуснати. Обща блок-схема за анализ на изображения на обекти включваща всички етапи на обработката е представена на Фиг. 1.1.



ФИГУРА 1.1: Обща блок-схема за анализ на изображения на обекти.

Първата обработка която се извършва на изображението, след като то бива заснето от камерата, се извършва от блока **Детектиране**. Тази обработка се извършва от т.нар. детектор, и целта е да се установи групата пиксели на изображението, които съставляват обекта на интерес. Вместо самата група от пиксели, правоъгълникът, който я загражда също се използва на практика. Самият процес на детектиране е сложен и сам по себе си, може да бъде описан със същата блокова схема, но без блока **Детектиране**. Само в най-елементарните случаи детекторът може да бъде изграден на база евристики или аналитични описания. В повечето случаи се използва обучаваща извадка, която, в зависимост от задачата, може да детектира заграждащите правоъгълници на обектите или групи от пиксели, които ги съставляват.

Следващият етап от обработката е блока **Нормализация**. Този блок в общия случай не е задължителен и обработката, която се извършва е свързана с нормализация на данните. Нормализацията може да бъде от различно естество и също така може да бъдат няколко от тях. Най-често срещаната нормализация е по отношение на осветлението. Вариациите в осветлението, както беше споменато, представляват един от основните проблеми при разпознаването на обекти.

Нормализацията в повечето задачи е от изключителна важност, но тя не се прилага в случаите, когато извадката е голяма и съдържа изображения на обекти с различна вариация на осветлението.

Блокът **Регистрация** служи за нормализация по отношение на координатната система в която се намират обектите. По-точно, цели се, позата и размера на изображенията на обектите (след детектирането), да бъдат подравнени по отношение на приетата координатна система. По този начин значително се намаляват вариациите, които възникват поради различен мащаб, поза или положение на обекта в сцената, което значително улеснява последващата работа. В някои случаи регистрацията може да се извърши с аналитични методи, но в общия случай тази задача се решава, като се използват алгоритми с обучение.

Следващият блок от обработката е **Извличане на признаци**, който е от основна важност. Извличането на признаци е обработка, която цели да извлече съществената информация от изображението на обекта спрямо задачата която се решава. Това, също така довежда и до намаляване на размерността на задачата, което от друга страна, ускорява последващите изчисления. Извличането на признаци, само по себе си, може да се извърши с аналитични средства или като се използва обучаваща извадка. Често извличането на признаци може да се интерпретира като избиране на признаци, защото алгоритмите избират най-информативните за целта признаци. Тази обработка е в основната на този труд и е разгледана подробно в следващата точка.

В зависимост от задачата, извлечените признаци се използват за класификация или за регресия. Обработката в блокът **Класификация** цели изображението на обекта да се класифицира в един от няколко предварително зададени класа (групи) $\gamma \subset \mathbb{N}$. Най-простият пример е класификация на изображения на човешки лица според пола, тоест $|\gamma| = 2$. Класификаторите в съвременното компютърно зрение се обучават, като се използва обучаваща извадка, с предварително аотирани от хора обекти.

Блокът **Регресия** е алтернативен на блока **Класификация** и в зависимост от задачата, на изхода си може да извежда реално число или изображение. В първия случай се цели да се извлече някаква количествена информация за обекта, който се анализира. Например, на база на изображение на човешко лице, да се определи неговата възраст. Пример, при който на изхода се извежда изображение може да бъде автоматичното състаряване или подмладяване на лица на индивиди. В този случай, на входа се подава изображение на човешко лице а на изхода се извежда същото лице, но състарено или подмладено. Регресията също изисква обучаваща извадка, която е предварително аотирани от човек.

1.2 Признакови пространства използвани при анализ на изображения на обекти и лица

Човешкото лице, като обект се явява от огромен интерес за компютърното зрение поради различните приложения в които то играе съществена роля. Важността му отдавна е разпозната от науката и възприятието и интерпретацията му се изучават за да се разбере какви са процесите които го формират [7], каква важност има за формирането на човешката мисъл и какво значение има за социалното

взаимодействие между хората [8]. Една от първите и значителни научни работи, свързани с анализа на човешки лица, е на Чарлз Дарвин [9] в която се търсят зависимости между: емоционалните състояния на хората и техните изражения, биологичните процеси и израженията, зависимостите между отделните изражения. Тази работа цитира научни трудове датиращи много преди нея, което само по себе си показва, че хората много отдавна са разпознали важността на човешкото лице като съществен обект от човешката природа. Основополагащ резултат от неврологията, свързан с текущата дисертация, е изследването направено от [10]. Авторите са изследвали центрове на мозъка, свързани с възприятието на лицата като са направили важните изводи, че детектирането на човешкото лице, не е непременно холистичен процес, а е достатъчно да бъдат показани само отделни части от него. По точно, формирането на възприятието е свързано с работата на много, но отделни неврони (или малки групи от неврони), всеки един от които, е специално адаптиран да разпознава отделни признаци на лицето (средно всяка група неврони е адаптирана до три признака). В резултат на това, групата от неврони адаптирани към отделни признаци на лицето образуват пространството на лицето. Това показва една важна насока на работа, свързана с автоматизирания анализ на лица а именно, че фокуса трябва да пада върху локалната, признаковата оценка на лицето. Но сам по себе си, локалният аспект не е достатъчен и също така е необходима глобална, холистична оценка на лицето, защото в крайна сметка тези групи от неврони принадлежат към пространството на лицето.

Въпреки, че научните области свързани с анализа на лица, са богати на резултати и са достигнали до важни изводи, които могат да се използват от компютърните науки, кодирането на информацията на лицето от компютъра така, че да постига същите възможности като човешкия мозък, все още остава нерешена задача. Съвременните изследвания показват, че човешката система за визуално възприятие съдържа целенасочени механизми които са фокусирани върху три основни типа модели на възприятие: холистично, компонентно и конфигурално [11]. Всеки един от тези модели може да бъде използван самостоятелно [12], [13], но и често те се считат за допълващи се [14], [15]. **Холистичното** възприятие моделира лицето като едно цяло, без изолирани части. **Компонентното** възприятие предполага, че лицевите компоненти се обработват отделно от човека. **Конфигуралното** възприятие моделира пространствените отношения между лицевите компоненти. Базирайки се на тези типове възприятия, в компютърните науки се въвеждат различни видове представяне на лицето, като избора на представяне зависи от контекста в който се разглежда. Например, представянето може да се раздели на няколко типа въз основа на пространствената информация, която се съдържа в него. Най-често срещаните такива са представяне на база на формата му или на база на външния му вид. Формата включва геометричните характеристики на лицето, докато външният му вид включва информацията, която се съдържа в текстурата на лицето. Друг вид представяне е на база изцяло пространствени или пространствено-времеви характеристики. Всички тези класификации на представянето на лицето са свързани с конкретна обработка на двумерно или тримерно изображение на лице. Извлечената информация при обработката е под формата на признаци и като резултат, лицето се представя в признаково пространство, което съдържа само съществената информация необходима за задачата, която се

решава. Поради това, построяването и класификацията на признаковите пространства е от съществено значение при анализирането на изображения на човешки лица от компютърните системи. Процеса по представяне на лицето в признаково пространство се нарича **извличане на признаци**.

Общата дефиниция за пространство в математически смисъл е: множество от елементи, което е снабдено със структура [16]. Структурата на това множество определя операциите, които могат да се извършват с елементите му и именно това го дефинира като пространство. Структурата с която е снабдено може да бъде например топологична, метрична, векторна, многообразие и т.н. Нека е дадено двумерно или тримерно изображение на човешко лице $X \in \mathcal{X}$, което принадлежи на класа от всички изображения на лица \mathcal{X} . Формално, извличането на признаци се описва математически с оператора $\Phi : \mathcal{X} \rightarrow \mathcal{F}$, който трансформира лицето в друго пространство \mathcal{F} , наречено признаково. Всяко признаково пространство има асоцииран с него математически оператор Φ , като по този начин се дефинира наредената двойка $\{\Phi, \mathcal{F}\}$ с която еднозначно се определя процеса по извличането на признаци.

Дефиниция. Признаково пространство на изображение (двумерно или тримерно) на човешко лице $\{\Phi, \mathcal{F}\}$ - Множество от елементи, снабдени с определена структура, които са генерирани като резултат от обработка, зададена с оператора $\Phi : \mathcal{X} \rightarrow \mathcal{F}$, действащ върху пространството на лицата \mathcal{X} .

Основните признакови пространства, с които ще се работи в този труд са снабдени със структура от типове хилбертово пространство или многообразие. Първото е аналог на евклидовото пространство и е широко разпространено в инженерните науки, докато второто представлява множество от елементи, снабдено с атлас. Важно разширение на хилбертовото пространство е в случая, когато то е снабдено с вероятностна мярка, при което може да се определя вероятността за извличане на определен елемент от пространството.

Избора на признаково пространство е централен при анализа на лица и той става въз основа на характеристиките на всички компоненти, които участват в образуването на признаковото пространство, определени в дефиницията. Така избора се прави въз основа на следните подходи:

- Налично знание за типа на пространството на лицата;
- Избор на признаковото пространство спрямо свойствата на оператора, който се използва за извличане на признаци;
- Избор спрямо самият характер на признаковото пространство;

С първият подход се постига еднозначност при анализа на лица, защото цялата информация за пространството, в което лежат лицата е налична, което обаче на практика е невъзможно поради огромният брой съществуващи вариации. В практическите задачи обаче, липсата на точната информация се компенсира с допускане за типа на пространството. Най-честите допускания са пространството на лицата да бъде от тип многообразие или от тип вероятно пространство. Спрямо тези допускания се избира оператора, който извлича признаците.

Третият подход е най-рядко използван на практика, поради това, че се използва само когато има ограничения свързани с последващите етапи на обработката на информацията.

Вторият подход, при който изборът се прави спрямо свойствата на оператора, е най-широко използван. Причината е в това, че експлицитният избор на оператор задава определени свойства на извлечените признаци, които задават признаковото пространство, което от своя страна отразява желанието на изследователя да извлече конкретен тип информация от изображението на лицето.

1.2.1 Класификация на признаковите пространства

За да могат да бъдат използвани на практика, признаковите пространства трябва да се охарактеризират с техните свойства и въз основа на това да бъдат класифицирани (групирани). Класификацията не може да бъде еднозначно определена поради големият им брой свойства, които често се припокриват. Въпреки това, използвайки основните им свойства е възможно да се постигне класифициране, което е достатъчно удобно за практически цели. Един от основните типове класифициране на признаковите пространства е свързано с определянето на параметрите на оператора. Ако параметрите са предварително зададени и фиксирани, то извличането на признаците става **без обучение (без учител)**, но ако параметрите се определят като функция на самите данни, то тогава извличането на признаци е с **обучение (с учител)**. Фиксираните параметри обикновено се определят аналитично чрез допускане на определени свойства на оператора. Пример за оператор с фиксирани параметри е преобразуването на Фурие при което коефициентите на базисните функции са фиксирани и предварително зададени, тоест, по никакъв начин не зависят от самите данни. Пример за вторият случай при който параметрите се определят от данните е алгоритъма Principal Component Analysis (**PCA**), при който параметрите зависят от статистическите зависимости между отделните елементи на данните.

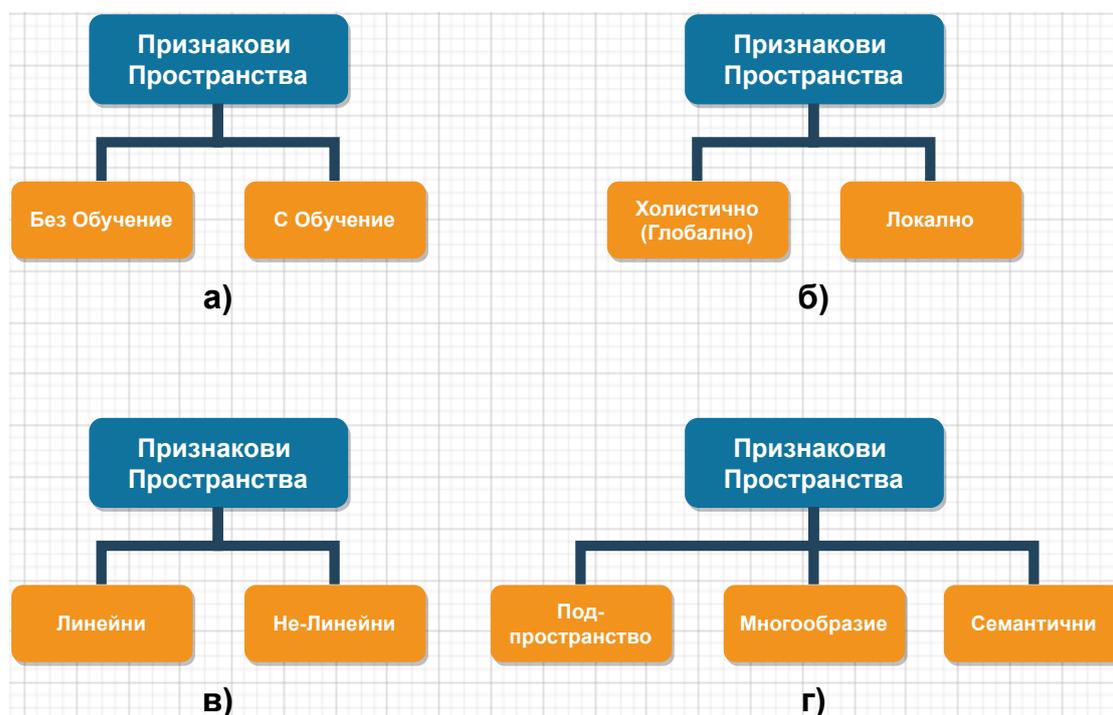
Друг тип основно класифициране на признаковите пространства е въз основа на областта на действие на оператора спрямо областта на изображението на лицето. Ако областта на действие на оператора е цялата област на изображението, тогава извличането на признаци се нарича **холистично (глобално)**, а ако областта на действие е под-област на цялата област на изображението тогава се нарича **локално**. Признаковите пространства при този тип класификация се наричат съответно холистични и локални. В случая на локално признаково пространство, операторът Φ може да бъде дефиниран като композиция от два оператора $\Phi \circ \mathcal{L}$, където оператора \mathcal{L} първоначално избира под-област от изображението (например блок с размери 5×5 пиксела), върху която се прилага Φ .

В зависимост от типа на преобразуването което се извършва от оператора, извличането на признаци може да се класифицира, като линейна или нелинейна операция. Това от своя страна, без да се следва строга математическа формалност, позволява признаковите пространства да се класифицират като линейни и нелинейни.

Ако се използва математическата структура на признаковото пространство като критерии за класификация, то тогава може да се направи класификацията на: под-пространства, многообразия или семантични пространства. Първият тип пространства се явяват под-пространства на пространството на което принадлежат изображенията. Това означава, че е направено допускането изображенията да принадлежат на линейно векторно пространство и признаковото пространство

съответно е под-пространство на него. Вторият тип са най-често срещани на практика, защото многообразието е по-богат инструмент при описанието на данните. Поради неговата сложност за работа обаче, структурата му не винаги се използва в явен вид. Третият тип пространства, семантичните, все повече се използва в научните изследвания през последните години [17]. В семантичните пространства има въведена семантика, която задава семантично значение на всеки обект от пространството. Това означава, че всеки обект от пространството има някакво смислово значение за хората. Например, най-често срещаното семантично пространство е свързано със семантичната анотация на лицето на човека при което, на всяка област от лицето има зададено значение. Например, групата от пиксели на окоото има зададено семантично значение око, групата от пиксели на областта на носа задава семантично значение нос, и т.н.

Важно е да се отбележи, че класифицирането на признаковите пространства описани до тук, не следва стриктни математически формулировки, а само приетите в компютърното зрение класификации, чиято основна цел е оптимално извличане на признаците, спрямо задачата, която се решава или приложението. Често този избор е чисто субективен и се определя от изследователя. В литературата също така се срещат и други типове класификация на признаковите пространства, но те са частни (изолирани) случаи, и не са обект на този труд. Направените до тук класификации на признаковите пространства спрямо характеристиките на оператора са дадени на Фиг. 1.2.



ФИГУРА 1.2: Условно разделяне на признаковите пространства в компютърното зрение: а) в зависимост от използване на данните; б) в зависимост от приложението на оператора върху областта на изображението; г) в зависимост от типа на оператора; д) в зависимост от типа на изходното пространство.

Представените на фигурата структурни диаграми са в по-прост вид, състоящи се от само два дъщерни възела, защото структурната диаграма представлява пълно описание на признаковите пространства спрямо техните характеристики и свойства ще бъде неудобна и непрактична за визуализация. Поради това, в текущият труд представянето на признаковите пространства и техните основни свойства ще бъдат представени в табличен вид. В Таблица 1.1 са изложени голяма част от признаковите пространства предложени в научната литература, представени от операторите които ги генерират. Описанието на колоните на таблицата и техните стойности е както следва:

- Колона *Об.*, съкратено от Обучение, показва дали параметрите на оператора зависят от данните, тоест дали е с обучение или не.
- Колона *Действие* показва дали действието на оператора е локално или холистично.
- Колона *Трансф.* съкратено от Трансформация показва вида на трансформацията, която се прилага от оператора, като тя може да бъде линейна или нелинейна.
- Колона *Вх./Иzx.* съкратено от Входно/Изходно показва, какво е допускането за типа на входното и изходното пространство, формално дефинирано математически. Видовете могат да бъдат: Евклидово (Евкл.), Линейно Подпространство (ЛПП), Многообразие (Мн.), Вероятностно (Вер.), Семантично (Сем.).
- Колона *Предимства* съдържа основните предимства на операторите. Следните съкращения и техните описания са използвани:
 - Аналитичност (Аналитич.) - това предимство показва, че оператора е построен използвайки аналитични изрази. Такъв тип оператори са най-предпочитани, защото може да се разбере обхвата им на действие.
 - Мулти-резолюция - това свойство показва, че оператора работи като използва различни мащаби на изображението. По принцип то е част от аналитичния израз на оператора, но тук е употребено отделно, защото е важно свойство.
 - Дескриптивност - това предимство касае операторите, които са без обучение, и които в почти всички случаи имат аналитично описание. То показва, че оператора има свойството да обхваща голяма част от вариациите на данните.
 - Вероятностни допускания за данните (Вер.) - това показва, че е направено допускане за случайни величини на данните, извлечени от вероятностно разпределение. В по-голяма част от случаите, допускането за вероятностното разпределение на данните е в аналитичен вид.
 - Голям капацитет (Кап.) - това предимство касае операторите които са с обучение. То показва, че оператора може да бъде обучен с данни, които имат голяма характеристична вариация или грубо казано, оператора може да опише сложни взаимовръзки между данните.

- Колона *Слож.* съкратено от Сложност съдържа изчислителната сложността на оператора, като тя е зададено условно като Голяма, Средна или Малка. Условната оценка на сложността се приема спрямо уейвлетното преобразуване, чиято сложност се счита за малка. Целта на тази условна оценка е да даде представа за изчислителната сложност на оператора, спрямо останалите представени в таблицата. Сложността определя само приложението на оператора и не касае изчисляването на параметрите в случаите, когато операторът е с обучение.

В научната литература има публикувани много други признакови пространства, които се генерират чрез използването на оператори, които са производни на операторите, представени в таблицата. Обемът на тези изследвания е огромен, и пълно описание на производните оператори е извън целите на текущият труд.

Групиране на операторите може да се забележи и от таблицата като всяка група се състои от няколко реда от таблицата, разделени от другите с тройна хоризонтална линия. Първата група включва уейвлетното преобразуване, Gabor и т.н. Тази група съдържа методи, които могат да се класифицират като класически, които използват техники от функционалния анализ за създаване на пространства с ортогонални или пре-определени базиси. Исторически погледнато, те са едни от първите методи които се използват на практика за разпознаването на обекти [18]. Втората група включва LBP, LTP и т.н. Тази група се появява по-късно във времето, като операторите са дефинирани, за да адресират недостатъците от първата група и/или да използват статистически зависимости на данните. Формално тези оператори са нелинейни и изходните им пространства рядко могат да бъдат определени като линейни. Третата група включва PCA, NMF, ICA и т.н., които използват статистически характеристики на данните, и обикновено са холистични методи. Появата на PCA също датира от времето на уейвлетите, но развитието на групата започва по-късно. Следващата група от оператори включва LPP, OLPP, LLE и т.н. Тази група от оператори също действа холистично, но развитието ѝ започва, след като са се установили недостатъците на предходната група. По точно, операторите в тази група използват допускания за принадлежност на данните към многообразия, което позволява да се използват техники от диференциалната геометрия. Развитието на тези оператори е започнало, когато се е установило, че оператори като PCA, ICA, LDA не могат напълно да моделират зависимостта между данните, базирайки се изцяло на допускането за вероятно пространство.

По сходен начин, в Таблица 1.2 са представени признакови пространства, като операторите които ги генерират са реализирани с дълбоки невронни мрежи. Те отново са групирани като всяка група е отделена от останалите с три разделителни хоризонтални линии. Първата група включва операторите VGG, AlexNet и т.н. Те са базирани на конволюционни невронни мрежи (Convolutional Neural Networks - CNN), но архитектурите им концептуално се различават, като всяка има своите предимства и недостатъци. Втората група включва операторите реализирани с т.нар. Transformers архитектури. Тези архитектури се отличават концептуално от CNN и се характеризират с въвеждането на т.нар. механизъм за внимание (attention). В повечето задачи свързани с анализ на двумерни изображения те се представят значително по-добре от CNN, но съответно изискват по-голяма изчислителна мощ и по-голяма обучаваща извадка. Третата група се характеризира с

това, че те работят с данни, моделирани с графи. Графите се характеризират с неравномерно дискретизиране на данните, както и задаване на свързаност между възлите на графа в явен вид. Поради това те са изключително подходящи за обработка на 3D изображения на лица, където всеки възел на графа представлява вектор в 3D пространството. Последната група оператори започна развитието си съвсем наскоро с идеите за семантична комуникация като основна част от 6-та генерация мобилни мрежи (6G) [17]. Реално тези оператори са базирани на дълбоки невронни мрежи, защото е необходим голям капацитет на обучение, но поради това, че генерират семантично пространство са отделени в самостоятелна група.

В следващите под-точки се дава дефиниция на всяка една от групите, въвеждат се основни понятия и се прави анализ на литературата.

1.2.2 Признакови пространства от тип линейни под-пространства и нелинейни пространства

Линейните признакови пространства се генерират, като се допуска, че пространството на лицата е под-пространство на евклидовото пространство, тоест $\mathcal{X} \subseteq \mathbb{R}^N$ и оператора Φ е линеен. Един оператор е линеен, когато за $\forall x, y \in \mathcal{X}$ следва:

$$\Phi(\alpha x + \beta y) = \alpha \Phi x + \beta \Phi y \quad (1.1)$$

където $\alpha, \beta \in \mathbb{R}$ са скалари. Тоест, линейната комбинация на елементи от пространството е равна на линейната комбинация на елементите в изходното пространство. Допускането, че пространството на изображенията на лица \mathcal{X} (и по принцип пространството на изображенията) е евклидово пространство, не е коректно, защото сумата на две лица представени като вектори от това пространство, не генерира нов вектор, който да е изображение на лице. Въпреки това, допускането за евклидово пространство е широко разпространено, защото в този случай оператора Φ представлява матрица, а прилагането му е умножение на матрица по вектор (или матрица). Тази операция е желана на практика, защото съвременните изчислителни машини я прилагат с висока скорост и ниска консумирана мощност. Тя се нарича умножение и акумулиране (Multiply and accumulate - MAC), като транзисторните схеми реализиращи процесорите я прилагат ефективно и при голяма тактова честота. Нещо повече, линейните оператори са основен елемент от съвременните невронни мрежи, което показва тяхната ефективност, въпреки некоректното допускане за пространството на изображенията лица. В много случаи, голяма част от задачите се приближават с линейни такива, именно за да могат да бъдат решени числено. Съществуват и линейни оператори които се построяват като се използва допускането, че пространството на лицата се описва с многообразие вградено в крайномерно евклидово пространство (LPP, OLPP, LLE). Този тип оператори, макар и линейни, не използват свойството (1.1).

Нелинейните признакови пространства се генерират от оператори които не изпълняват условието (1.1). Такива оператори например могат да бъдат комбинация от линейни оператори и нелинейни функции или изцяло нелинейни оператори. Примери за нелинейни оператори са морфологичните операции, които работят чрез използване на операциите за сравнение. Друг пример са операциите, свързани с избиране на минимални или максимални елементи от група от пиксели (LBP, LTP, SURF). Методите с ядрото са друг широко използван клас, при който всеки елемент от изображението на лицата се изобразява в пространство с по-висока размерност, където се извършват определен тип операции (KPCA, KLDA). Макар в основната на невронните мрежи да лежат линейни операции, те по същество са силно нелинейни, което се дължи на активиращите функции на невроните. Поради това, че нелинейните оператори предоставят по-големи възможности за обработка на информацията, те постигат по-добри резултати на практика спрямо линейните, но за сметка на голям брой изчисления, което, в зависимост от приложението, не винаги е постижимо на практика.

ТАБЛИЦА 1.1: Характеризация на признаковите пространства в зависимост от операторите които ги генерират.

Оператор	Об.	Действие	Трансф.	Вх./Изх.	Предимства	Слож.
Уейвлетно преобразуване [19]	Не	Локално	Линейна	Евкл./ЛПП	Аналитичност	Малка
Bio Inspired Features (BIF) [20]	Не	Локално	Линейна	Евкл./ЛПП	Аналитичност	Малка
Gabor [21]	Не	Локално	Линейна	Евкл./ЛПП	Мулти-рез. Аналитичност	Малка
Log-Gabor [22], [23]	Не	Локално	Линейна	Евкл./ЛПП	Мулти-рез. Аналитичност	Малка
Local Binary Patterns (LBP) [24]	Не	Локално	Нелинейна	Евкл./Мн.	Дескриптивност	Голяма
Local Ternary Patterns (LTP) [25]	Не	Локално	Нелинейна	Евкл./Мн.	Дескриптивност	Голяма
Histogram of Oriented Gradients (HOG) [26]	Не	Локално	Нелинейна	Евкл./Мн.	Дескриптивност	Голяма
Scale Invariant Features Transform (SIFT) [27]	Не	Локално	Нелинейна	Евкл./Мн.	Дескриптивност	Голяма
Speed-Up Robust Features (SURF) [28]	Не	Локално	Нелинейна	Евкл./Мн.	Дескриптивност	Средна
Local Phase Quantisation (LPQ) [29]	Не	Локално	Нелинейна	Евкл./Мн.	Дескриптивност	Средна
Part Based SIFT (PB SIFT) [30]	Не	Локално	Нелинейна	Евкл./Мн.	Дескриптивност	Голяма
Quantized Local Zernike Moments (QLZM) [31]	Не	Локално	Нелинейна	Евкл./Мн.	Дескриптивност	Голяма
Dense Bag Of Words (BoW) [32]	Не	Локално	Нелинейна	Евкл./Мн.	Дескриптивност	Голяма
Sparse Coding [33]	Не	Локално	Нелинейна	Евкл./Мн.	Дескриптивност	Голяма
DAISY [34]	Не	Локално	Нелинейна	Евкл./Мн.	Дескриптивност	Голяма
Weber Local Descriptors (WLD) [35]	Не	Локално	Нелинейна	Евкл./Мн.	Дескриптивност	Голяма
Self-Similarity Descriptor (SSD) [36]	Не	Локално	Нелинейна	Евкл./Мн.	Дескриптивност	Голяма
Principal Component Analysis (PCA) [37]	Да	Холистично	Линейна	Вер./ЛПП	Аналитичност, Вер. п-во.	Малка
Linear Discriminant Analysis (LDA) [38]	Да	Холистично	Линейна	Вер./ЛПП	Аналитичност, Вер. п-во.	Малка
Independent Component Analysis (ICA) [39]	Да	Холистично	Линейна	Вер./ЛПП	Вер. п-во	Малка
Non-negative Matrix Factorization (NMF) [40]	Да	Холистично	Линейна	Вер./ЛПП	Аналитичност	Малка
Kernel PCA (KPCA) [41]	Да	Холистично	Нелинейна	Вер./Мн.	Аналитичност, Капацитет	Голяма
Kernel LDA (KLDA) [42]	Да	Холистично	Нелинейна	Вер./Мн.	Аналитичност, Капацитет	Голяма
Generalized Discriminant Analysis (GDA) [43]	Да	Холистично	Нелинейна	Вер./Мн.	Вер. п-во., Капацитет	Голяма
Locally Linear Embedding (LLE) [44]	Да	Холистично	Линейна	Мн./ЛПП	Аналитичност, Капацитет	Малка
Laplacian Eigenmaps (LE) [45]	Да	Холистично	Линейна	Мн./ЛПП	Аналитичност	Малка
Locality Preserving Projections (LPP) [46]	Да	Холистично	Линейна	Мн./ЛПП	Аналитичност	Малка
Kernel LPP (KLPP) [46]	Да	Холистично	Нелинейна	Мн./Мн.	Аналитичност, Капацитет	Малка
Orthogonal LPP (OLPP) [47]	Да	Холистично	Линейна	Мн./ЛПП	Аналитичност, Капацитет	Малка
Locally Discriminant Embedding (LDE) [48]	Да	Холистично	Линейна	Мн./ЛПП	Вер. п-во, Капацитет	Малка
Spectral Regression (SR) [49]	Да	Холистично	Линейна	Мн./ЛПП	Аналитичност, Капацитет	Малка
Neighborhood preserving projections (NPP) [50]	Да	Холистично	Линейна	Мн./ЛПП	Аналитичност, Капацитет	Малка
Conformal Embedding Analysis (CEA) [51]	Да	Холистично	Нелинейна	Мн./Мн.	Аналитичност, Капацитет	Голяма

ТАБЛИЦА 1.2: Характеризация на признаковите пространства в зависимост от операторите реализирани с дълбоки невронни мрежи.

Оператор	Об.	Действие	Трансф.	Вх./Изх.	Предимства	Слож.
AlexNet [52]	Да	Холистично	Нелинейна	Мн./Мн.	Капацитет	Голяма
VGG-16 [53]	Да	Холистично	Нелинейна	Мн./Мн.	Капацитет	Голяма
GoogLeNet [54]	Да	Холистично	Нелинейна	Мн./Мн.	Капацитет	Голяма
ResNet [55]	Да	Холистично	Нелинейна	Мн./Мн.	Капацитет	Голяма
YOLO [56]	Да	Холистично (Локално)	Нелинейна	Мн./Мн.	Капацитет	Голяма
ViT [57]	Да	Холистично	Нелинейна	Мн./Мн.	Капацитет	Голяма
VOLO [58]	Да	Холистично	Нелинейна	Мн./Мн.	Капацитет	Голяма
SWIN v2 [59]	Да	Холистично	Нелинейна	Мн./Мн.	Капацитет	Голяма
BViT [60]	Да	Холистично	Нелинейна	Мн./Мн.	Капацитет	Голяма
Conviformers [61]	Да	Холистично	Нелинейна	Мн./Мн.	Капацитет	Голяма
ViTDet [62]	Да	Холистично	Нелинейна	Мн./Мн.	Капацитет	Голяма
СОМА [63]	Да	Холистично	Нелинейна	Мн./Мн.	Капацитет	Голяма
DynGCNN [64]	Да	Холистично	Нелинейна	Мн./Мн.	Капацитет	Голяма
MeshCNN [65]	Да	Холистично	Нелинейна	Мн./Мн.	Капацитет	Голяма
FML [66]	Да	Холистично	Нелинейна	Мн./Мн.	Капацитет	Голяма
GraphCMR [67]	Да	Холистично	Нелинейна	Мн./Мн.	Капацитет	Голяма
DeepSC [68]	Да	Холистично	Нелинейна	Мн./Мн.	Капацитет	Голяма
RL-based adaptive semantic coding [69]	Да	Холистично	Нелинейна	Мн./Мн.	Капацитет	Голяма
U-net [70]	Да	Холистично	Нелинейна	Мн./Мн.	Капацитет	Голяма
Submanifold Sparse CNN [71]	Да	Холистично	Нелинейна	Мн./Мн.	Капацитет	Голяма
Segmenter [72]	Да	Холистично	Нелинейна	Мн./Мн.	Капацитет	Голяма

1.2.3 Намаляване на размерността на признаковите пространства

В предните раздели беше споменато, че с извличането на признаци се цели да се извлече информацията, съществена за задачата, която се решава. Намаляването на размерността на данните е процес, при който признаковото пространство е с по-ниска размерност и той може да се счита, като метод за извличане на съществена информация. Но друга по-важна характеристика на намаляването на размерността е свързана с т.нар. „проклятието на размерността” (curse of dimensionality) [73]. Това е ефект който, при нарастването на размерността на данните разгледани като вектори от евклидово пространство, експоненциално нараства необходимият брой от тях за да може надеждно да се определи вероятностното разпределение от което са извлечени. Така при нарастването на размерността на данните, те изглеждат все по-разредени в пространството и в този си смисъл са все по-малко представителни при определяне на вероятностното разпределение. Клъстеризацията в пространства с висока размерност става по-трудна, защото всички вектори започват да изглеждат на равни разстояния един от друг. Това лесно може да се провери като се изчисли отношението между максимално и минимално разстояние, като функция на размерността на пространството и се вземе границата при $d \rightarrow \infty$. Построяването на разделяща повърхнина от класификатора става все по-трудна задача с нарастване на размерността [74]. Поради това, съществуват много методи, прилагани преди класификатора, с които да бъде намалена размерността. Линейните статистически преобразувания и техните разширения реализирани с ядро са най-широкия клас от методи, използвани в практиката.

Намаляването на размерността има и други предимства, като намаляване на изчислителна сложност на последващите обработки или предаване на информация на разстояние. В този труд, намаляването на размерността, като метод за извличане на признаци се използва, за да се избегне ефекта „проклятието на размерността”.

Разбира се, намаляването на размерността, макар да се използва широко не е задължителен метод. Има алтернативи, при които се извличат признаци, чиито признакови пространства са дори с по-висока размерност от пространството на изображенията и представляват пре-определени матрици. В общия случай, обаче, последващата обработка е свързана с приложение на методи с разредени коефициенти което може да се счита като имплицитно намаляване на размерността (избор на признаци).

1.2.4 Признакови пространства генерирани от линейни статистически преобразувания

Линейните статистически преобразувания спадат към класа оператори, които са линейни, тоест генерират линейни под-пространства, както е описано в т. 1.2.2. Също така се допуска, че данните принадлежат към дадено вероятностно пространство и поради това използват статистически методи за изчисляване на коефициентите на операторите. Този тип методи е един от първите в компютърното зрение, които доказват важността на извличането на признаци с оператори, които се обучават от данни.

Първият метод използван в литература е PCA [37]. В този основополагащ труд, авторите решават задачата за разпознаване на лица. Допускането е, че изображенията на лицата са извадка от вероятно многомерно гаусово разпределение. На база на това допускане авторите използват PCA за да извлекат признаци като проектират данните в линейно под-пространство което те наричат „пространство на лицата”. Самият линеен оператор който се генерира от PCA запазва основната част от вариациите в данните и премахва тези които са незначителни. По този начин се премахва шум, както и вариации които се смятат за незначителни за задачата. Самият праг с който се решава кои вариации да бъдат премахнати се избира експериментално. Премахването на компонентите които отговарят на малките вариации в крайна сметка резултира в намаляване на размерността на данните. След като бъдат проектирани изображенията в пространството на лицата се извършва класификация с евклидово разстояние, за разпознаване на идентичността на неизвестното лице. Те също така демонстрират използваемостта на метода за детектиране на лица. Основното предимство на PCA е премахването на излишни вариации в данните, като по този начин се постига избягване на проблема с „проклятието на размерността” и поради това този метод се използва широко. Недостатък на метода е допускането, че данните са извлечени от гаусово разпределение което в повечето случаи е невярно и често води до незадоволителни резултати.

Друг метод който се базира на статистически линейни трансформации е Linear Discriminant Analysis (LDA) предложен в [38]. Авторите изтъкват недостатъка на метода PCA, че не взема предвид информацията за класовете на обектите когато задачата е за класификация. По-точно, при изчисляването на PCA се максимизира, както между-класовите, така и вътрешно-класовите вариации, което не е ефективно за класификацията. Изхождайки от това, те предлагат да се максимизират между-класовите вариации, като съвременно се минимизират вътрешно-класовите вариации. Първо се построява матрицата на между-класовите вариации S_B , като се изчислява ковариационната матрица на средните стойности на всеки клас. Аналогично се построява матрицата на вътрешно-класовите S_W вариации като сума от ковариационните матрици на данните от всеки клас. Използвайки тези матрици се решава обобщената задача за собствените стойности. Матрицата от собствените вектори на решението се използва за да се проектират данните в признаковото пространство, което е под-пространство на евклидовото, и което те наричат Fisherfaces. Характерна особеност на признаковото пространство е, че размерността му е равна на броя на класовете минус едно. Авторите демонстрират превъзходството на метода спрямо PCA в задачата за разпознаване на лица при различни осветености. Въпреки недостатъците на PCA, той се използва като предварителна стъпка за премахване на шум. Основен недостатък на метода LDA е, че се използват ковариациите на данните което означава, че се използват статистически зависимости само от втори ред, с което се опростява задачата и съответно води до намаляване на ефективността (точността на разпознаване).

Използване на статистически зависимости на данните за построяване на линеен оператор е предложено в [39] и наименуван Independent Component Analysis (ICA). Характерно за този метод е, че не се прави допускане за вероятно разпределение, от което са извлечени данните. Вместо това в методът се търси линеен оператор, при който трансформираните данни да са статистически независими в

признаковото пространство. Това е силен критерии който се използва за да се изолират вариации в изображението, които не са релевантни за задачата. Също така, това позволява да се извлекат признаци, които са по-описателни. Това води до използването на статистически зависимости от по-висок ред, което е предимство спрямо методите, разгледани в тази точка. От гледна точка на честотният анализ, който се извършва с Фурие базис, това означава, че освен честотната информация, фазовата такава също се използва в модела, която е дори от по-голямо значение от честотната, защото съдържа информация за структурата на обектите [39]. Този алгоритъм за разлика от разгледаните до тук, не намалява размерността на признаковото пространство. За сравнение с PCA, ICA алгоритъмът предложен в тази работа не прави допускание за гаусово разпределение на данните, а за супергаусово такова. Матрицата на линейната трансформация се намира чрез решаване на оптимизационна задача която се стреми да максимизира взаимната информация между входните и признаковите данни. Това се постига като целевата функция се задава да бъде ентропията на признаците, като функция от матрицата на линейната трансформация. Авторите решават задачата за разпознаване на лица, предлагайки две различни архитектури. При първата, всяко изображение на лице се разглежда, като случаен вектор, а при втората, всеки пиксел се разглежда като случайна величина. В резултат, първата архитектура намира базиси, които са холистични, сходно на PCA, а втората локални, сходно на уейвлет функциите. Авторите демонстрират ефективността на алгоритъма, като показват, че той превъзхожда PCA. Предимството на този тип линейна статистическа трансформация е, че използва статистики от по-висок ред което означава, че могат да се моделират по-сложни процеси. Недостатък на алгоритъма е необходимостта от голям брой изчисления и дори често, в зависимост от типа данни, алгоритъмът може да не намери решение.

Методът Non-negative Matrix Factorization (NMF) е предложен от [40]. Този метод може да се счита за сходен на ICA, но основната разлика е, че има ограничение коефициентите които образуват базиса на оператора да са не-отрицателни. Методът, също така, не използва информация за анотация на изображенията. Поради това, че елементите на оператора са ограничени да са не-отрицателни, генерираните базиси са силно разредени, съдържащи много нулеви коефициенти което от другата страна означава, че те са пространствено групирани. Поради този факт, самото извличане на признаци може да се каже, че се извършва по части, съставлящи изображението. Така например в случая с анализа на лица, NMF алгоритъма създава базис, който се концентрира около частите на лицето, които съдържат много пространствени вариации, като очите, носа, устата. Нека $\mathbf{X} \in \mathbb{R}_+^{F \times L}$ е матрица, сформирана от изображения на обектите, като всяко изображение представлява една колона от матрицата. Задачата, която решава NMF е $\mathbf{X} \approx \mathbf{Z}\mathbf{H}$, където $\mathbf{Z} \in \mathbb{R}_+^{F \times M}$ е матрицата, която съдържа базисните вектори, а $\mathbf{H} \in \mathbb{R}_+^{M \times L}$ съдържа коефициентите на линейните комбинации на \mathbf{Z} , които реконструират изображенията. Тоест, при дадено изображение \mathbf{x}_i извличането на признаци се извършва, като $y_i = \mathbf{Z}\mathbf{x}_i$. Решението се намира като се решава оптимизационна задача за минимизиране на Kullback–Leibler (KL) дивергенцията между \mathbf{X} и $\mathbf{Z}\mathbf{H}$ при ограничения за положителност на коефициентите на \mathbf{Z} и \mathbf{H} . Авторите предлагат да разширят метода, така че да бъде подходящ за задачите за класификация на обекти, като включат между-класовите и вътрешно-класовите вариации подобно

на метода LDA, но на вече трансформирани данни. Методът е тестван за задачата за разпознаване на лицеви изражения и резултатите от експериментите показват ефективността на алгоритъма спрямо останалите алгоритми. Основното му предимство е, че могат да се интерпретират базисните функции и съответно да се модифицират спрямо задачата. Недостатък на метода е, че факторизацията е реализирана с приближение, което означава, че не е гарантирано достигането на минимум който отговаря на оригиналната задача.

1.2.5 Признакови пространства от тип многообразие

Многообразието е математически обект, който позволява по-точно описание на данните или операторите. Използва се много често при анализа на обекти поради факта, че изображенията са с голяма размерност, а самите обекти могат да бъдат описани с данни с по-ниска размерност. Анализирани в тази под-точка методи също така се наричат методи с Обучение с многообразие. Дефиницията на многообразие е както следва:

Дефиниция Многообразието \mathcal{M} е хаусдорфово пространство с изброим базис от отворени множества със свойството всяка точка от \mathcal{M} има околност, която е хомеоморфна на отворено подмножество на \mathbb{R}^n [75].

Диференцируемостта на многообразието е едно от желаните свойства, което позволява да се решават оптимизационни задачи. Не винаги обаче, аналитичните свойства на многообразието могат да се използват на практика поради необходимия огромен брой изчисления. Параметрите на операторите, които се използват за генериране на този тип признакови пространства се изчисляват от данните и затова спадат към класа оператори с обучение.

Често при определяне на параметрите на операторите се допуска те да принадлежат на дадено многообразие. Например, многообразието $SO(3)$ се състои от всички матрици с размерност 3×3 , които са ортонормални и които се използват при калибрация на камери или определяне на позата на обектите [76]. В този случай често се търси явен аналитичен вид за работа с многообразието. Когато се търси модел на данните чрез многообразие обаче, самият му характер не може да бъде зададен аналитично поради липсата на знание за процеса на формирането му (в някои по-прости задачи това не е така). Тогава, задачата е да се открие характеристичната структура на данните, а именно признаковото пространство, която е вградена в пространство със значително по-ниска размерност от тази на данните. По-конкретно се допуска, че данните $x_1, \dots, x_k \in \mathcal{M}$ а \mathcal{M} е вградено в \mathbb{R}^n и задачата се свежда до алгоритмично решение, което намира съответстващи (представителни) признаци $y_i \in \mathbb{R}^d$ така, че $d \ll n$.

Методът Locally Linear Embedding (LLE) [44] е един от първите методи които демонстрират ефективността на този тип методи. Този метод вгражда данните, разгледани като вектори, в пространство с ниска размерност, като намира оптималната линейна реконструкция на околност на вектора. Първо, за всяка точка x_i от данните се търсят най-близките k съседа с алгоритъм за най-близкия съсед k -NN [77], като k се избира експериментално. След това се търсят тегла W_{ij} , която линейна комбинация да представя оптимално (в смисъла на най-малки квадрати) точката x_i от k -те си съседа. След като са намерени теглата, се решава сходна

задача, при която теглата са фиксирани, а се търсят проекции на точките в признаковото пространство, което е с по-ниска размерност, така че да удовлетворяват условието за линейна комбинация на най-близките k съседа, но в признаковото пространство. Обобщено, може да се каже, че методът, за всяка точка от данните, апроксимира околността на точката, като (локално) евклидово пространство с базиси W_{ij} и след което се търсят проекции с по-ниска размерност, които да удовлетворяват условието за локално евклидово пространство определено от W_{ij} . Методът демонстрира, как изображения на различни изражения на лицето на един човек, могат да бъдат проектирани в двумерната равнина без да има припокриване между проектираните вектори за всяко изражение. Предимството на метода е неговата простота за изпълнение. Недостатък е, че при голям обем от данни необходимият брой изчисления може да нарасне значително. Друг недостатък е, че построяването на локалния базис е в смисъл на най-малки квадрати което го прави чувствителен към невалидни данни.

Модификация на LLE е предложена в [50] и наименувана Neighborhood Preserving Projections (NPP). Разликата се състои в това, че вместо да се търси апроксимация на околността на точка от данните с линеен базис, се търси базис който се изчислява като се използва LPP. По този начин се гарантира по-точна апроксимация на околността като се вземе предвид, че околността на точката също е част от многообразието. Авторите демонстрират метода, като използват синтетични данни, и го сравняват с LLE. Предимството, което се вижда от експерименталните резултати е, че методът успява да постигне по-ниска размерност на пространството, в което се вграждат данните. Основен недостатък на този метод е нуждата от голям брой изчисления, защото за всяка точка трябва да се изчислява обобщената задачата за собствените стойности.

Един от основополагащите методи свързан с моделирането на данните чрез многообразие е Laplacian Eigenmaps (LE) [45]. В основата му стои използването на оператора на Laplace-Beltrami, който гарантира оптимално вграждане за многообразието. Този оператор, обаче, е дефиниран за многообразия M с непрекъснатата структура в смисъла на гладки и компактни многообразия. При анализа на данните, обаче, извадките са краен брой, поради което се налага приближаване с дискретни структури. Така многообразието M , на което принадлежат данните се се приближава с граф, който се нарича граф на близост, ръбовете на който се определят чрез определено правило за близост между данните. Едно такова, например, може да е зададена ϵ околност около x_i и всички възли на графа, които влизат в тази околност се свързват с ръбове. Допълнително, всеки ръб се асоциира с тегло което може да е евклидовото разстояние или например, нелинейна функция от него. След като е построен графа, апроксимиращ многообразието, се преминава към приближение (апроксимация) на оператора Laplace-Beltrami, което се нарича graph-Laplacian. Приближаването се извършва, като се използват теглата на ръбовете на графа, които се явяват коефициенти на ядрото на Laplace-Beltrami, чрез което ядро се моделира разпространението на топлината. Вграждането на данните в пространството с по-ниска размерност се извършва, като се използват собствените вектори на обобщената задача за собствените стойности, използваща graph-Laplacian. Този тип методи също така се наричат „Спектрална клъстеризация”, при които обаче, апроксимацията на Laplace-Beltrami се извършва по сходен, но различен начин. Авторите демонстрират ефективността на метода LE върху

синтетични данни, текст и говор. Основното предимство на метода е това, че се базира на солидна теория свързана с оператора за вграждане в Laplace-Beltrami. Недостатък на метода е липсата на еднозначност при построяването на графа приближаващ многообразието. Друг недостатък е нуждата от разлагане на собствени вектори и стойности, които при големи обеми от данни може да изисква огромен брой изчисления (квадратична зависимост от данните), независимо че в голяма част от случаите, матриците са разреждени.

Друг метод използващ предположението признаците да бъдат многообразие вградено в пространство с по-ниска размерност е Locality Preserving Projections (LPP) [46]. Този метод, подобно на LE построява граф приближаващ многообразието, но вместо да решава обобщената задача за собствените стойности, се решава линейна апроксимация на него, като се използва допускането, че данните се вграждат в признаковото пространство чрез линейна трансформация. Това свежда проблема до решаването на частно на Релей, което се решава чрез задачата за собствени вектори. Авторите демонстрират работата на метода с разпознаване на лица и клъстеризация на лица според позата и сравняват работата му със сходни методи. Недостатък на метода е нуждата от разлагане на матрици, но това може да се избегне като предварително се избере броят на желаните собствени вектори които да се изчислят. Друг недостатък е допускането, че вграждането в признаковото пространство, което е с по-ниска размерност се извършва чрез линейна трансформация.

Подобрение на LPP е предложено в [47] и наименувано Orthogonal LPP (OLPP). Както името подсказва, линейното преобразуване, с което се извършва вграждането при този методи е с ортогонален базис. Разликата от LPP се състои в това, че локалната апроксимация на евклидовото пространство изисква базисът да е ортогонален. Решението се намира, като се използват множители на Лагранж и се построява функцията на Лагранж, която след диференциране се свежда до итеративна процедура за намиране на собствените вектори и стойности. Като предварителна стъпка авторите също така въвеждат изчисляването на PCA с цел да се намалят изчисленията. Те демонстрират превъзходството на метода като го сравняват с LPP и PCA за задачата за разпознаване на лица. Макар методът да превъзхожда сходните методи, основен недостатък остава линейната трансформация, която вгражда данните, макар и тя да бъде ортогонална. Друг недостатък е невъзможността за използване на метода на ядрото за получаване на нелинеен оператор.

Метод който използва вграждането на многообразие, използвайки и дискриминативност е предложено от [48], наименувано Locally Discriminant Embedding (LDE). Този метод се използва в приложенията за класификация на обекти в които има зададени от човек етикети. Разликата с останалите методи споменати в тази подточка е, че тук се построяват две приближения на многообразието G и G' като се взимат предвид етикетите. Графът G се построява, като между всеки два възела се слага ръб ако те принадлежат на същият клас, а при G' се слага ръб ако те не принадлежат на един и същи клас. Така построените графи генерират две различни graph-Laplacian апроксимации L и L' , като с тях се решава обобщената задача за собствени стойности, като L се ползва от дясната страна на равенството, а L' от лявата. Авторите демонстрират превъзходството на алгоритъма като го сравняват с PCA, LPP и други методи за задачата за разпознаване на лица.

Недостатък на алгоритъма е квадратичното нарастване на изчисленията спрямо обема от данни както и допускането, че вграждането се извършва чрез линейна трансформация подобно на LPP.

По-нов подход за намиране на LE, наречен COLES е предложен в [78]. Методът използва контрастно обучение (contrastive learning) за намиране на коефициентите на оператора. За разлика от LE, COLES използва избрана граф-конволюционна невронна мрежа с която се постига вграждане в признаковото пространство, което е с по-ниска размерност. Методът използва извличане на случайни данни, които не принадлежат на разпределението на данните, като част от целевата функция на оптимизационната задача. Авторите също така демонстрират, как предложената методология от тях може да се използва и за LPP. Те доказват превъзходството на метода, като го сравняват с други методи за задачата за класификация на текстови документи.

1.2.6 Обща рамка за моделиране на линейни статистически преобразувания

Методите за построяване на линейни статистически преобразувания описани в т. 1.2.2 и методите за генериране на признакови пространства от тип многообразие описани в т. 1.2.5 следват общ модел на построяване при който крайната цел в общия случай е намаляване на размерността на пространството на изображенията. Това позволява създаването на обща рамка за работа с тях, която е предложена в работата [79]. Рамката касае операторите от тип матрици и тензори които проектират евклидово пространство в под-пространство на него. Също така, в нея влизат методите които използват зависимости между данните въз основа на които се изчислява базис чрез разлагане по собствени стойности: PCA, LDA, LPP, LE, LDE, както и техните производни. Рамката използва моделирането на зависимостите между данните като граф, след което цели вграждането на този граф в пространство с по-ниска размерност, запазвайки структурата му (graph embedding).

Нека са дадени N изображения представени, като вектори $x_i \in \mathbb{R}^n$ и нека бъдат наредени в матрица $X = [x_1, x_2, \dots, x_N]$. Рамката използва като основа ненасоченият граф $G = \{X, W\}$, където матрицата $W \in \mathbb{R}^{N \times N}$ съдържа тегла с които се описват зависимостите между всеки два елемента от данните и играе централна роля. Най-простият пример е теглата да се изчисляват, като евклидовото разстояние между всеки две изображения, но в общият случай се допускат и отрицателни тегла. На база на тегловната матрица се изчислява graph-Laplacian по следният начин:

$$L = D - W, \quad D_{ii} = \sum_{i \neq j} W_{ij}, \quad \forall i \quad (1.2)$$

Рамката допълнително въвежда и наказателен граф $G = \{X, W^p\}$, при който тегловната матрица $W^p \in \mathbb{R}^{N \times N}$ съдържа тегла определящи зависимости между данните, които трябва да бъдат подтиснати. Грубо казано, тази матрица се използва за да се контролира въздействието на W . Нека в общият случай $y_i = F(x_i)$, където при линейните преобразувания $F(\cdot)$ е матрица. За определянето на $F(\cdot)$ рамката решава следната оптимизационна задача:

$$y^* = \arg \min_{y^T B y = d} y^T L y \quad (1.3)$$

Коефициентът d е константа, а матрицата B представлява ограниченията на оптимизационната задача, за да се избегнат тривиални решения (в случая нулеви решения). Тази матрица по принцип е диагонална, но може да бъде и матрицата на Лаплас от наказателният граф: $B = L^p = D^p - W^p$. Така дефинирана рамката позволява дефиниране на различни задачи чрез подходящ избор на G и потенциално G^p . Важно за този труд е да се подчертае, че макар рамката да не допуска вероятности разпределения на данните тя позволява да се моделират статистически зависимости от втори и по-висок ред чрез подходящите G и G^p , което от своя страна я прави подходяща като основа за описание на линейните статистически преобразувания. Друго важно предимство на рамката важно за този труд е това, че тя позволява постигане на нелинейни зависимости чрез метода на ядрото описан в следващата точка.

Авторите на рамката предлагат и нов метод наречен Marginal Fisher Analysis (MFA) който стъпва на нея. Този метод е подходящ за приложение в задачите за класификация, където се използват анотирани данни. Те подчертават недостатъкът на LDA който е допускането, че данните принадлежащи на всеки клас са извлечени от гаусово разпределение, което е рядкост на практика. Тоест, статистическите зависимости от втори ред използвани в LDA в общият случай са ограничаващи. Те предлагат два критерия, които да бъдат взети предвид, а именно вътрешно-класовата компактност и между-класовата разделимост. Първият се моделира, като се построява графът G , а теглата на W се дефинират чрез алгоритъмът k -NN приложен на ниво клас. Вторият критерий се задава чрез наказателният граф G^p , теглата на който се изчисляват на ниво между-класова разделимост, определена от алгоритъма k -NN. Те демонстрират превъзходството на MFA над PCA, LDA и LPP при разпознаване на изображения на лица върху няколко бази данни. Предимството на MFA са използваните по-коректни критерии за построяване на вътрешно и между класовите зависимости, което има и по-ясна интуитивна интерпретация. Недостатък е използването на k -NN както и нуждата за разлагане на собствени стойности, което за големи бази данни може да се окаже невъзможно.

1.3 Методи в разпознаването на изображения на обекти, реализирани с ядро

Изчисления в пространства с висока размерност и потенциално безкрайномерни са непосилни за съвременните компютри. Това ограничение на изчислителните системи възпрепятства използването на голям брой методи които могат да предоставят по-големи възможности за разпознаването на обекти. Въпреки това, ако се допуска че алгоритмите работят в хилбертово пространство, има възможност чрез т.нар. метод с ядрото (kernel trick), скаларните произведения да бъдат лесно изчислени, имплицитно в пространства с висока размерност и дори безкрайномерни (на теория). Това позволява на методите за машинно обучение и по-специално линейните преобразувания да бъдат лесно реализирани в пространства с висока размерност. Това позволява не само постигане на нелинейност, а и увеличаване на капацитета на обучение на алгоритмите.

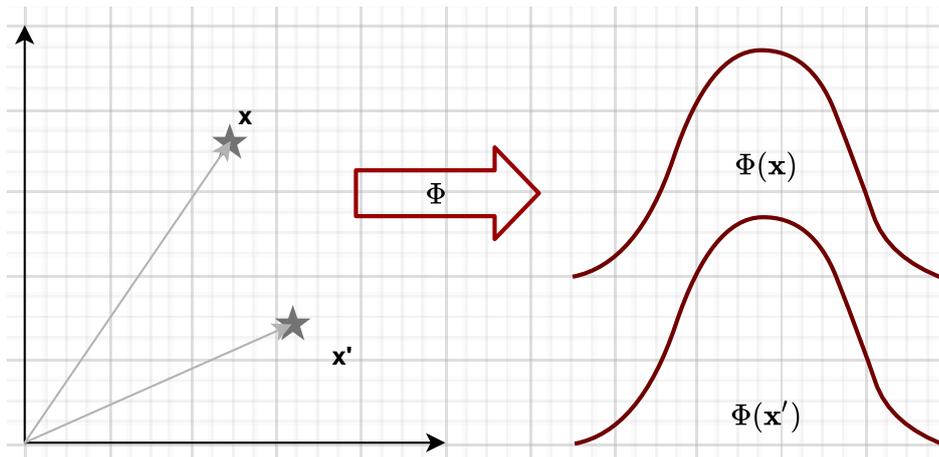
1.3.1 Имплицитна реализация на нелинейност чрез методът с ядрото

В основата на метода стои т.нар. ядро (kernel), което представлява функция на два аргумента приемаща стойности в реалните числа $k(\cdot, \cdot) \mapsto \mathbb{R}$. Функцията не може бъде произволна и трябва да отговаря на определени свойства, но в основата си тя трябва да измерва сходство между аргументите. Използвайки тази функция имплицитно се извършват скаларни произведения в пространство, което е с по-висока размерност от пространството на изображенията.

Нека \mathcal{X} е пространството на изображенията (изображенията на лицата). Преобразуването от това пространство в пространство с по-висока размерност се извършва с преобразуването [80]:

$$\phi : x \in \mathcal{X} \mapsto \phi(x) \in \mathcal{H} \quad (1.4)$$

където \mathcal{H} е хилбертово пространство със съответното скалярно произведение. Схематично за две изображения това е представено на Фиг. 1.3.



ФИГУРА 1.3: Визуализация на изображението Φ от пространството на лицата в хилбертово пространство образувано от гаусови функции

От фигурата се вижда как вектори от двумерната област се изобразяват като гаусови функции чрез Φ . Това преобразуване позволява да се работи с функции, които предоставят повече възможности и в частност са безкрайномерни (на теория).

Функцията $k(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ се нарича ядро, ако съществува хилбертово пространство \mathcal{H} над реалните числа и изображение $\phi : \mathcal{X} \mapsto \mathcal{H}$, такова че $\forall x, y \in \mathcal{X}$ следва $k(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}$. За да може ядрото да се използва, е необходимо то да е възпроизвеждащо за хилбертовото пространство. Дефиницията следва, като се използва оценъчният функционал на Дирак:

Дефиниция 1.3.1. Нека функциите в $f \in \mathcal{H}$ приемат стойности в \mathbb{R} . За $x \in \mathcal{X}$ оценъчният функционал на Дирак в точката x се задава, като $\delta_x : \mathcal{H} \mapsto \mathbb{R}$ и дефинира $\delta_x(f) = f(x)$.

Използвайки оценъчният функционал на Дирак се дефинира Хилбертово пространство с възпроизвеждащо ядро по следният начин:

Дефиниция 1.3.2. Хилбертовото пространство \mathcal{H} е Хилбертово пространство с възпроизвеждащо ядро когато оценъчните функционали на Дирак са ограничени отгоре, тоест $\exists M > 0, M < \infty$ такава, че $\|\delta_x(f)\| < M\|f\|_{\mathcal{H}}, \forall x \in \mathcal{X}, \forall f \in \mathcal{H}$.

Използвайки теоремата за представянето на Riesz [80], за $\forall f \in \mathcal{H}$ и $\forall x \in \mathcal{X}$ може да се намери функцията $k_x \in \mathcal{H}$, такава че:

$$f(x) = \delta_x(f) = \langle f, k_x \rangle_{\mathcal{H}} \quad (1.5)$$

Тъй като k_x е функция, принадлежаща на \mathcal{H} , следователно може да се намери $y \in \mathcal{X}$, такава че:

$$k_x(y) = \delta_y(k_x) = \langle k_x, k_y \rangle_{\mathcal{H}} \quad (1.6)$$

Това позволява да се дефинира ядрото $k(x, y) = \langle k_x, k_y \rangle_{\mathcal{H}}$. Понеже k е ядро, тоест $k_x = \phi(x)$, следва $k(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}$, което именно е възпроизвеждащото свойство и означава, че ядрото възпроизвежда скаларното произведение в хилбертовото пространство. Това се нарича и методът с ядрото (kernel trick) понеже избягва изчисляването на скаларно произведение в \mathcal{H} , а директно те се изчисляват имплицитно с ядрото k .

Тази конструкция изхожда от знанието за функциите ϕ , които изобразяват пространството на изображенията в \mathcal{H} . На практика, при конструирането на ядрото, не винаги тези функции са известни. Поради това се използва следната теорема на Moore-Aronszajn:

Теорема 1.3.1. За всяко положително определено ядро $k : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$, съществува единствено хилбертово пространство с функции над \mathcal{X} , за което, ядрото е възпроизвеждащо.

Тази теорема позволява всяка функция на два аргумента със стойности в реални числа, за която може да се докаже, че е положително определена, да се използва в практическите задачи като ядро.

Основното свойство на хилбертовите пространства с възпроизвеждащо ядро, което позволява намирането на решения в това пространство, е свързано с теоремата за представителя.

Теорема 1.3.2. Нека е дадено положително определено ядро $k : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$, асоциирано с възпроизвеждащо хилбертово пространство \mathcal{H} . Нека също така са дадени:

- Обучаваща извадка $(x_1, y_1), (x_2, y_2), \dots, (x_M, y_M) \in \mathcal{X} \times \mathbb{R}$;
- Строго монотонно нарастваща функция $g : [0, \infty) \mapsto \mathbb{R}$;
- Функция, измерваща грешката $E : (\mathcal{X} \times \mathbb{R}^2)^M \mapsto \mathbb{R} \cup \{\infty\}$;

Тогава, всеки минимизатор на емпиричният риск:

$$f^* = \arg \min_{f \in \mathcal{H}} \{E((x_1, y_1, f(x_1)), (x_2, y_2, f(x_2)), \dots, (x_M, y_M, f(x_M))) + g(\|f\|)\}$$

приема следното представяне:

$$f^*(\cdot) = \sum_{i=1}^M \alpha_i k(\cdot, x_i),$$

където $\alpha_i \in \mathbb{R}$.

Накратко казано, теоремата за представителя казва, че решението на функционали, които са регуляризирани в пространства с висока размерност или безкрайномерни, лежи в крайномерно пространство което е линейната обвивка от данните, преобразувани в хилбертовото пространство. Това значително намалява нуждата от огромен брой изчисления и представя елегантно решение на задачите в машинното обучение, където винаги се минимизира емпиричният риск и в общият случай решението, което се търси, бива регуляризирано.

Методът с ядрото е основно приложим, когато има скаларни произведения. Те предимно се прилагат при линейните преобразувания, и в частност статистическите линейни преобразувания. Процедурата, по която се прилага метода е сравнително проста. Първо се избира ядрото в зависимост от задачата, която се решава. След това с ядрото се построява матрицата на Грам, така че за всеки елемент ij съответният коефициент се изчислява като $k_{ij} = k(x_i, y_j)$. Така построената матрица се замества в алгоритъма вместо скаларните произведения между x_i и x_j . Основна задача при тези методи остава избора на ядро, както и неговите параметри. Този избор не е очевиден и зависи от характера на данните и от задачата която се решава.

1.3.2 Линейни статистически трансформации, реализирани с ядро, използвани в разпознаването на изображения на обекти

Едно от първите приложения за разпознаване на обекти използвайки метода с ядрото и РСА е предложен в [41]. Авторите предлагат Kernel PCA (КРСА), метод който е сходен на оригиналният РСА, но първо изображенията се преобразуват в пространство с по-висока размерност. Авторите използват полиномниално ядро от ред d , за да приложат метода с ядрото и да построят матрицата на Грам. След това се изчисляват собствените стойности и аналогично на РСА, изображенията се проектират в пространство с по-ниска размерност, използвайки теоремата за представителя. Те демонстрират ефективността на предложения алгоритъм за разпознаване на лица, използвайки линейната версия на класификатора Support Vector Machines (SVM). Резултатите са сравнени с РСА, LDA, конволюционна невронна мрежа, невронна мрежа с вероятно вземане на решение, и показват, че предложеният от тях алгоритъм постига по-малка грешка при разпознаването. Основният недостатък на алгоритъма е нуждата от изчисляване на собствени вектори, което расте квадратично с нарастване на данните, което при големи извадки става невъзможно.

Версия на метода LDA реализирана с ядро и наречена Kernel LDA (KLDA) е предложена в [42]. Авторите подхождат аналогично както при КРСА, но вместо една матрица на Грам, те построяват две използвайки гаусово ядро, едната описваща вътрешно-класовите вариации а другата между-класовите вариации. Решават

обобщената задача за собствените стойности и намира собствените вектори чрез теоремата за представителя. Демонстрират ефективността сравнявайки го с SVM и AdaBoost класификаторите, използвайки синтетични данни. Основен недостатък на алгоритъма също се явява нуждата от голяма изчислителна мощ при големи извадки.

Сходен алгоритъм наименуван Kernel Direct Discriminant Analysis (KDDA) е предложен в [81]. В основата на този метод стои direct LDA, при който се търсят решения в сечението на нулевите решения на матриците на вътрешно и между-класовите вариации. Предложена е версия на direct LDA, базирана на метода с ядрото чрез използването на гаусово ядро. Демонстрирана е ефективността на метода, като е сравнен с KPCA и Generalized Discriminant Analysis (GDA) за задачите за разпределение на лицата в двумерното пространство и задачата за разпознаване на лица при която използват k-NN класификатор. Основното предимство на метода е, че работи много по-добре от останалите за задачи които са с малка обучаваща извадка което се дължи на постигната нелинейност и direct LDA. Като недостатък на метода може да се посочи голямата алгоритмична сложност и нуждата от изчислителна мощ с нарастването на данните.

Версия на LPP реализирана с метода с ядрото наименувана Kernel LPP (KLPP) е предложена в [46]. Авторите демонстрират как решенията на KLPP са сходни с тези на LPP, но вместо линейна проекция се използват решенията от теоремата на представителя.

Версия с метода с ядрото на MFA е предложена в [79]. Използвано е гаусово ядро за построяването на матрицата на Грам. Предложеният метод е приложен за разпознаването на лица и демонстрират, че той превъзхожда стандартните линейни методи, както и MFA. Предимствата на метода са, че използва нелинейността, реализирана чрез метода с ядрото, както и зависимостите на данните, които вече се изразяват в хилбертовото пространство. Недостатък е зависимостта на изчисленията от броя на данните.

1.4 Изводи

От направеният обзор на методите за извличане на признаци, и по конкретно, линейните статистически преобразувания и тяхната модификация която е реализирана с ядро за приложение на разпознаване на изображения на обекти, могат да бъдат направени следните изводи:

1. Най-важният от етапите в системите за разпознаване на изображения на обекти е извличането на признаци, защото решава проблеми, като: отстраняване на шум, „проклятието на размерността”, намалява размерността и извлича информация която е важна за решаваната задача.
2. Извличането на признаци чрез използването на оператори с фиксирани коефициенти не винаги дава добри резултати. Необходимо да има адаптивност към данните чрез използване на статистическите зависимости между тях. В тази връзка линейните статистически преобразувания представляват интерес, и също така заради лесната им приложимост.

3. Използването на статистически зависимости от втори ред не е достатъчно за да се моделират процесите на формиране на изображението, а е необходим по висок ред. Поради това, описанието на данните като многообразия се явява универсален метод, с който могат да се постигне по-точно моделиране на изображенията на естествени сцени.
4. Апроксимирането на многообразия с графи и построяването на graph-Laplacian е един от успешните методи с които може да се постигне добро приближение на многообразието. Методите с използването на graph-Laplacian позволяват да се моделират както линейните статистически преобразувания, които използват зависимости от втори ред, така и зависимости от по-висок ред. Това ги нарежда като универсален метод, който е конструиран на база на геометрични съображения, които също така обхващат и статистическите такива.
5. Методът с ядрото позволява лесно да се постигне нелинейност на вече съществуващи линейни статистически преобразувания, като имплицитно ги преобразува в пространство с по-висока размерност. Това преобразуване е за сметка на не-голямо нарастване на изчислителната сложност, но предоставя по-голям капацитет на обучение. Основен недостатък на този подход е нуждата от голям брой изчисления във фазата на обучение при налични големи обучаващи извадки, като в някои случаи това може да се окаже невъзможно.

Като обобщение на изводите може да се каже, че линейните статистически преобразувания моделирани чрез graph-Laplacian и реализирани с ядро са мощен инструмент при разпознаването на изображения на обекти. Въпреки това проблемите, които трябва да се адресират при този подход са свързани предимно с изчислителната сложност, която нараства квадратично с броя на данните.

1.5 Дефиниране на целта и основните задачи на настоящата дисертация

Целта на настоящата дисертация е разработване на методи и алгоритми за разпознаване на изображения на лица, базирани на линейните статистически преобразувания, техните реализации по метода с ядрото, и използване на метода с ядрото в етапа на класификация.

От поставената цел на дисертационния труд произтичат и следните *основни задачи*:

1. Разработване на алгоритми за детектиране и регистрация на човешки лица, представени в 2D и 3D изображения .
2. Разработване на алгоритъм за определяне на възраст в групи (разпознаване на възраст), чрез вграждане на възрастовото многообразие в подпространство с линейно преобразуване и нелинейно такова, реализирано по метода с ядрото.
3. Разработване на алгоритъм за разпознаване на изражения на лица, базиран на локални признаци и избор на най-информативните от тях, които запазват вграждането на многообразието на данните в подпространство.

4. Разработване на алгоритми за разпознаване на изражения на лица на база геометрични признаци, използвайки метода с ядрото обучен с големи обеми от изображения, заснети в неконтролирани условия.
5. Разработване на алгоритъм за разпознаване на 3D изображения на лица, с извличане на локални геометрични признаци и проектирането им в дискриминативно подпространство.
6. Разработване на алгоритъм за разпознаване на 3D изображения на лица, с допълващи се локални геометрични признаци, комбинирани чрез множество ядра за увеличаване капацитета на обучение.
7. Функционална проверка на действието на разработените алгоритми, чрез програмна реализация и стандартни бази данни, за оценка на тяхната ефективност.

2 Разработване на алгоритъм за определяне на възраст от изображения на лица

Съвременните тенденции на развитие на информационните технологии налагат нуждата от подобряване на взаимодействието между хората и компютрите с цел, да се повиши достъпността на компютърните системи. Като част от тези усилия, много изследователи са работили в областта на автоматичен анализ на изображенията на лица, например при изработването на интелигентен интерфейс човек-компютър. В този контекст са разработени системи които могат да идентифицират лица, да разпознават емоции, пол и възраст. Системите за автоматична възрастова оценка биха били изключително полезни в различни приложения на взаимодействието с компютрите. В резюме, тези приложения могат да се обобщят до следните:

- Взаимодействието на човек с компютър: Ако компютърът може да определи възрастта на потребителя, операционната система може да бъде адаптирана спрямо възрастта на потребителя. Такава система също може да се използва в комбинация с контрол на Интернет достъп до сайтове с неподходящи материали за непълнолетни лица и за автоматично откриване на сайтове с порнографско съдържание;
- Автоматичната възрастова оценка може да се използва за извличане на лицевите изображения от бази данни, които са свързани с възрастта;
- Възрастта може да бъде генерирана изкуствено, за да се постигне състаряване, което може да се използва например във филмовата индустрия;
- Приложения и в полицията за по-лесно разкриване на извършител на престъпление. Например, лицето извършител е заснето от охранителни камери и чрез системата се оценява неговата възраст и по този начин може по-лесно да бъде открит в базата данни или да се определи неговата социална среда.

Определянето на възрастта, освен като задача сама по себе си, е и предизвикателство в подобни задачи за разпознаване на лица [82], [83]. Поведението на хората и тяхното предпочитание са различни в различните възрасти [84], което показва големия потенциал на приложения на автоматичната възрастова оценка. В сравнение с други промени на лицето, ефекта на стареенето показва три уникални характеристики:

- Напредъка на стареенето е неконтролируем и никой не може да предотврати стареенето. Тази процедура е бавна и необратима;

- Различните хора остаряват по различни начини, което води до модели на персонализираното стареене. Моделът на стареене на всеки човек се определя, както от неговите/нейните гени, така и от много външни фактори, като например: здраве, начин на живот, околни условия и др.;
- Моделите на стареене са времеви данни и процесът на стареене се подчинява на реда на времето. Външният вид на лицето, по отношение на бръчки и нарушена текстура ще се отрази на всички стари лица, но няма да окаже влияние върху тези на младите индивиди.

Всяка от тези характеристики допринася за трудностите на автоматичната възрастова оценка най-вече, защото хората не могат свободно да контролират промяната при стареене. Това прави събирането на достатъчно данни за обучение на алгоритми за автоматична оценка на възрастта изключително трудно. Въпреки това, с напредъка на научните работи в областта, има достатъчно бази данни които могат да се използват за разработката на такъв тип алгоритми.

В текущата глава е предложен алгоритъм за оценка възрастта на база на изображения на лица, която класифицира лицето в няколко възрастови групи, който алгоритъм е базиран на линейни статистически преобразувания и техните реализации с ядро. Предимството на предложеният алгоритъм, спрямо другите в литературата е използването на *метод за вграждане на многообразието на данните в подпространство* (възрастовото многообразие), запазвайки геометрията им, както и *увеличаване дискриминативността*.

2.1 Анализ на литературните данни по определяне на възраст от изображения на лица

Добре известно е, че изображенията на човешки лица предоставят възможност да се определят човешки признаци като идентичност, раса, пол и други. Определянето на човешката възраст е много важна тема на изследователска дейност в приложения като интерфейс човек-компютър, 2D и 3D разпознаване на лица и виртуална реалност. Най-често срещаните методи за определяне на човешката възраст се базират на анализ на изображения на лица [85]. Те могат да бъдат най-общо класифицирани в четири главни категории: антропометрични модели [82], подпространства на възрастови модели [86], възрастови многообразия [87] и регресия [88]. Първата категория, антропометричните модели, измерват геометрични съотношения на лицето и бръчките на кожата, на база на които класифицират лицето в няколко възрастови групи. Тези методи са подходящи за грубо определяне на възрастта, основно за ранна възраст. Методите за определяне на възраст чрез възрастови многообразия, подпространства на възрастови модели и регресия, използват обучаващи извадки на лица с аотирана възраст. Тези подходи превъзхождат антропометричните модели по отношение на постигната точност, и това определя тяхната популярност в научните среди. При методите с многообразия се цели моделиране на процеса на стареене чрез многообразие, което се вгражда в пространство със значително по-ниска размерност от пространството на изображенията [87]. Моделите с регресия също моделират процеса на стареене като многообразие, или по-простият вариант подпространство, но при тях фокуса

пада върху регресията, поради което те могат да се разглеждат като отделен клас методи.

Антропометрията на лицето е наука за измерване на размерите и пропорциите на човешки лица. Farkas [89] дефинира антропометрията на лицето чрез измервания на 57 лицеви ориентира (ключови точки) върху човешки лица на различни възрасти, от ранна детска възраст до зряла възраст. За определяне на растежа на хората, обикновено се използват съотношения на разстоянията между тези ориентира, вместо директно използване на математическите модели [90]. Има две основни причини хората да не използват математическата формула за оценка на възрастта: 1) Математическият модел не може да характеризира профила на главата по естествен начин, особено когато възрастта е близка до тази на възрастните, и 2) профилът на главата е труден за измерване от 2D изображения на лица.

Една от най-ранните научни работи в областта на възрастовата класификация от изображения на лицето, е трудът на Kwon и Lobo [82]. В своята работа те предлагат изчисляването на шест съотношения на разстоянията на лицеви компоненти и ги използват за класифициране в два класа: бебета и възрастни. Използвана е база данни, съставена от 47 лица, която не е публично достъпна. Резултатите от определянето на възрастта са докладвани само на база отделните отношения, а не като средна стойност върху базата данни. По-късният труд на [91] предлага използването на 8 съотношения за да моделират прогресирането на възрастта на лица, които са между 0 и 18 годишна възраст. Тяхната цел е да адаптират задачата за разпознаване на лица спрямо прогресирането на възрастта. Сходна работа е на Gunay и Nabiyeu [92], където също се предлага вариант, базиран на антропометрични характеристики.

Забележка: Всички статии за определяне на възраст, разгледани по-долу в тази точка, използват тестовата процедура leave-one-person-out (LOPO), освен ако не е упоменато друго. При тази процедура, тестовата извадка се състои от изображенията само на един човек (изображенията за всички възрасти на този човек), а обучаващата извадка включва всички останали изображения. По този начин се избягва обучение на алгоритъма с изображения на лица от същия индивид, с който ще бъде тестван алгоритъмът, и съответно, изместване на оценката му.

Една от първите успешни работи, които използват регресия, с която се моделира стареенето в линейни подпространства, е наименувана AGing pattErn Subspace (AGES) и е предложена в [93]. Този подход моделира индивидуалното стареене на лицето чрез подпространство. Възрастта при дадено входно изображение на лице се определя чрез проекцията му в подпространство, което най-добре може да възстанови изображението на лицето. Първоначално се извличат признаци използвайки статистически модели на лицето предложени в [94]. След което се построяват $p-1$ на брой признакови вектора, където p е броят на годините, които ще се определят. Всеки един от признаковите вектори се проектира в подпространството, което се построява с PCA (Aging Pattern). След което се реконструира изображението на лицето и се сравнява грешката от възстановяване. Признаковият вектор за който тази грешка е най-малка, определя възрастта на индивида. Авторите тестват алгоритъма, използвайки базата данни FG-NET [95]. Те сравняват ефективността му с други два алгоритъма, както и оценката извършена от хора, като демонстрират, че алгоритъма постига по-добра средна точност при определянето на възрастта. Основен недостатък на алгоритъма е допускането, че

процеса на стареене може да се моделира в подпространства, построени с PCA.

Друг метод използващ регресия е предложен в [96]. При формулировката на проблема е допуснато, че анотациите на възрастта могат да бъдат сгрешени, и на база на това, се формулира множество от регресии на които им се усреднява предсказването с цел да се намали ефекта от сгрешените анотации. Всяка една регресия е формулирана чрез метода с ядрото, но вместо едно ядро са използвани множество ядра. За определяне на възрастта се прилага линейна комбинация от ядрата, последвана от проекция. Обучението на алгоритъма се извършва като се решава положително полу-определена квадратична оптимизация. Авторите сравняват ефективността на алгоритъма с AGES, използвайки FG-NET базата данни, и демонстрират намаляване на средната грешка на определената възраст. Основен недостатък на алгоритъма е, че обучението се извършва чрез решаване на квадратична оптимизация, която може да стане невъзможна за изчисление при голяма извадка.

Друг алгоритъм, използващ регресия за определяне на възраст, е предложен в [97]. Авторите използват линейна регресия реализирана с ядро, за предсказване на възрастта. Извлечените признаци са базирани на дискретното косинусово преобразуване върху блокове с размери 8x8 пиксела, както и техните координати. Ядрото, което използват, се базира на Gaussian Mixture Models (GMM), които коефициенти се изчисляват при обучение. Квадратичната форма на ядрото се използва като мярка за сходство между отделните блокове. За тест на ефективността на предложеният алгоритъм се използва базата FG-NET и е сравнен с алгоритъма AGES, и този предложен в [96]. Демонстрирана е най-малка средна грешка на предсказаната възраст. Въпреки, че се справя по-добре от алтернативите, негов основен недостатък е високата изчислителна сложност, поради необходимостта от изчисление на теглата на GMM. Друг недостатък е липсата на обосновка за избора на дискретното косинусово преобразуване за извличане на признаци.

Авторите на [98] предлагат алгоритъм използващ Support Vector Regression (SVR) [74] с който да определят възрастта. Признаците, които се извличат, са комбинация между холистични и локални, като първите се базират на формата на лицето а вторите са LBP (виж т. 1.2.1). Холистичните признаци се използват за да се класифицира лицето в две възрастови групи: 0-20 години и 21-69 години. За всяка възрастова група се прилага регресия реализирана с SVR за определяне на възрастта. Оценка на ефективността на алгоритъма се прави използвайки базата данни FG-NET, но докладваното подобрене е в рамките на статистическата грешка.

Авторите на [99] предлагат Spatially Flexible Patch (SFP) за да определят възрастта. SFP представлява блок, който е част от изображението и кодира позицията му, както и коефициентите на дискретното косинусово преобразуване. След като се построят SFP за всяко изображение и за цялата обучаваща извадка, за всеки един от тях се построява GMM модел, който предсказва възрастта, но само за съответният блок. Възрастта се определя като максималното правдоподобие от сумата на предсказаните възрасти на всички SFP блокове. Авторите демонстрират ефективността на алгоритъма като го сравняват с [96] и многослоен перцептрон използвайки базата данни YAMANA, която се състои от 8000 изображения на женски лица с японски произход и възраст 0-93 години. Макар докладваната определена средна възраст да е по-добра от останалите алгоритми, алгоритъма не е

сравнен с популярната FG-NET база данни, поради което не може да се направи обективна оценка за работата му.

Поход, който първо класифицира лицата в две възрастови групи и след това прилага регресия за определяне на възрастта е предложен в [100]. Авторите извличат два типа признаци, холистични, базирани на Active Appearance Models (AAM), и локални, базирани на Uniform Local Ternary Patterns (ULTP). Преди извлечането на признаци се прилага преобразуването contourlet, но в логаритмичната област, с цел отстраняване на шум и подчертаване на бръчките на лицата. Върху извлечените признаци се прилага KLDA, с което се реализира намаляване на размерността до признаков вектор от 30 елемента. Този вектор се използва за да се класифицира лицето с SVM в две възрастови групи, 0-20 и 21-69 години. За всяка възрастова група се използва SVR за точното определяне на възрастта. Алгоритъмът е тестван върху базата данни FG-NET и въпреки, че резултатите за определяне на възраст по-добри от алгоритъма AGES, те са по-лоши от тези на алгоритъм, реализиран с извличане на BIF признаци.

В работата [101] авторите предлагат комбинация от холистични и локални признаци за определяне на възраст. Холистичните се състоят от AAM, а локалните от LBP, върху които се прилага на PCA за да се намали размерността. След това се използва Least Angle Regression (LAR), за да изберат най-информативните признаци измежду холистичните и локалните. Избраните признаци се подават на SVR, за да се определи възрастта. Алгоритъмът е тестван с базата данни UIVCPAL, но ефективността му не е сравнена с други алгоритми, поради което не може да се прецени ефективността му обективно.

Един от най-успешните алгоритми, използвайки извличане на признаци и регресия, е предложен в [102]. Първата стъпка от алгоритъма е извличане на областта на лицето използвайки Active Shape Models (ASM), като по този начин се премахва фона и косата. След това се извличат признаци с функции на Gabor с 16 скали и 8 ориентации. Това довежда до увеличаване на размерността и поради това, се прилага намаляване на размерността използвайки статистическите характеристики на признаците. Върху избраните признаци се прилага SVR за определяне на възрастта. Авторите тестват алгоритъма върху базите FG-NET и MORPH, и го сравняват с [96]. Те докладват значително по-ниска грешка при определянето на възрастта. Основен недостатък на алгоритъма е липсата на обективност при избора на статистическите методи за избор на признаци, както и сравнителен анализ с други методи за избор на признаци.

Алгоритъм, който използва многообразия, за да моделира процеса на стареене, е предложен в [87]. Първата стъпка от алгоритъма е детектиране на лицето и изрязването му за премахване на фона. Следващата стъпка е обработка на детектираните и подравнени лица чрез алгоритъма Conformal Embedding Analysis (CEA), който се използва за моделиране на многообразия и вграждането им в пространство с по-ниска размерност, тоест намаляване на размерността. При моделирането, авторите използват и методи с които се интегрира дискриминативност, за да се подобрят последващата класификация. Ръбовете на графите се построяват само ако има сходство между две изображения, което сходство се измерва чрез експоненциалната функция от корелацията между изображенията. След като е намалена размерността на данните се прилага многофакторна линейна регресия. Авторите

тестват ефективността на алгоритъма, използвайки базата данни UIUC-IFP, която не е достъпна публично. Алгоритъмът е сравнен с други методи за моделиране на многообразия, като NPR, LPP и OLPP (виж т. 1.2.1), и докладват по-добри резултати при определянето на възрастта. Алгоритъмът не е сравнен с други алгоритми за определяне на възраст, поради което, не е възможно обективно да се определи неговата ефективност.

Друг алгоритъм използващ моделирането на процеса на стареене с многообразие е предложен от [103]. Моделирането на лица чрез многообразие се извършва след като лицата бъдат детектирани и изрязани, с цел отстраняване на фона. Авторите използват LPP и OLPP за да моделират многообразието, като правят сравнителен анализ, и доказват че OLPP предоставя по-добри резултати. Извлечените признаци с OLPP се използват, за да се определи възрастта чрез регресия, която е модификация на SVR. Тази модификация прилага корекция при наличие на големи локални вариации в предсказаните резултати. Авторите използват FG-NET за оценка на ефективността на алгоритъма и докладват, че предложеният метод превъзхожда AGES и този предложен в [96]. Основен недостатък на алгоритъма е липсата на обективен подход при определяне на прозореца за оценка на локалните вариации.

Алгоритъм който моделира само геометричните характеристики на лицето с многообразие е предложен в [104]. Авторите използват само геометричните характеристики на лицето, които се описват от 27 ключови точки, детектирани с алгоритъма ASM. Всяка една от тези ключови точки е с две координати принадлежащи на координатната система на изображението. Използвайки факта, че детектираните точки на лицето са равнини в 27-мерното пространство, те ги трансформират в ортогонални матрици с алгоритъма Singular Value Decomposition (SVD), след което ги използват като точки от многообразие на Grassman. Използвайки експоненциални карти на многообразието на Grassman, те изчисляват близост между геометрията на лицата чрез геодезичното разстояние. Алгоритъма е тестван с базата данни FG-NET. Авторите докладват по-добри резултати от алгоритъмът AGES, което е доказателство, че геометрията на лицето съдържа голяма част от информацията свързана със стареенето. Основен недостатък на алгоритъма е липсата на обективност при избора на многообразието на Grassman .

2.1.1 Изводи от литературния анализ по определяне на възраст от изображения на лица

Изводите от анализа на литературата по определяне на възраст от изображения на лица могат да се обобщят до следните:

- Геометричните признаци на лицето (формата) са с по-голяма тежест при определянето на възраст, поради което е необходимо, те да бъдат взети предвид, дори и имплицитно;
- Точността на алгоритмите силно се влияе от регистрацията на изображенията, както и осветеността, поради което е необходимо да бъде използвана предварителна обработка, която да адресира тези проблеми;
- В случаите, когато не се използва многообразие, е подходящо използването на локални и холистични признаци.

- Класификацията в групи, като предварителна стъпка, подобрява определянето на възраст;
- Методите, които използват моделиране на възрастта чрез многообразие предоставят по-малка средна грешка;
- В случаите, когато признаците са с висока размерност е необходимо да се приложи PCA, за да се избегне „проклятието на размерността“.

2.2 Класификация по възрастови групи с линейни статистически преобразувания, реализирани по метода с ядрото

Определянето на възраст може да се извършва по групи, при което задачата от регресия, тоест определяне на години, се свежда до класификация в една от предварително определени възрастови групи. Този подход може се използва, например в криминалистиката, когато е необходимо да се определи пълнолетие на индивид от изображение на лицето му. Друг пример е класификацията в три или четири възрастови групи при провеждане на маркетингови изследвания. Предимството при класификация във възрастови групи, пред точното определяне на възраст, се състои в намаляване на очакваната грешка, поради намаляване на влиянието на индивидуалните вариации. Увеличава се и обучаващата извадка за всяка група, за сметка на точната възраст на индивида. Така задачата се опростява, което е и предпоставка за намаляване на изчислителната сложност.

Предложеният алгоритъм решава задачата за класифициране на изображение на лице в няколко, предварително зададени възрастови групи. Той е ново-предложена комбинация от няколко метода базирани на линейни преобразувания, всеки един адресира различен етап от обработката. Основното предимство на предложения алгоритъм е бързодействието, без да се прави компромис с ефективността му. Бързодействието е важна характеристика на алгоритмите, като благодарение на него може да се постигне работа в реално време, поради което то се взема предвид. Такива са, например, системите за вградени приложения които са с ограничен изчислителен ресурс и ниска консумация. Бързодействието на предложения алгоритъм се постига като в етапа на извличането на признаци, съответните линейни преобразувания се комбинират в едно. По отношение на ефективността, два фактора допринасят за нея. Първият е използването на линейно преобразуване, и/или неговата реализация с ядро, запазващо многообразието на данните при вграждането им в подпространство. Вторият фактор, който допринася за качеството, е използването на нелинейният класификатор SVM, реализиран с ядро. Този тип класификатори се характеризират с ниско бързодействие, но в конкретният случай, бързодействието се постига благодарения на намалената размерност, постигната при извличането на признаци.

2.2.1 Блок-схема на потока на информацията и анализ на предложения алгоритъм

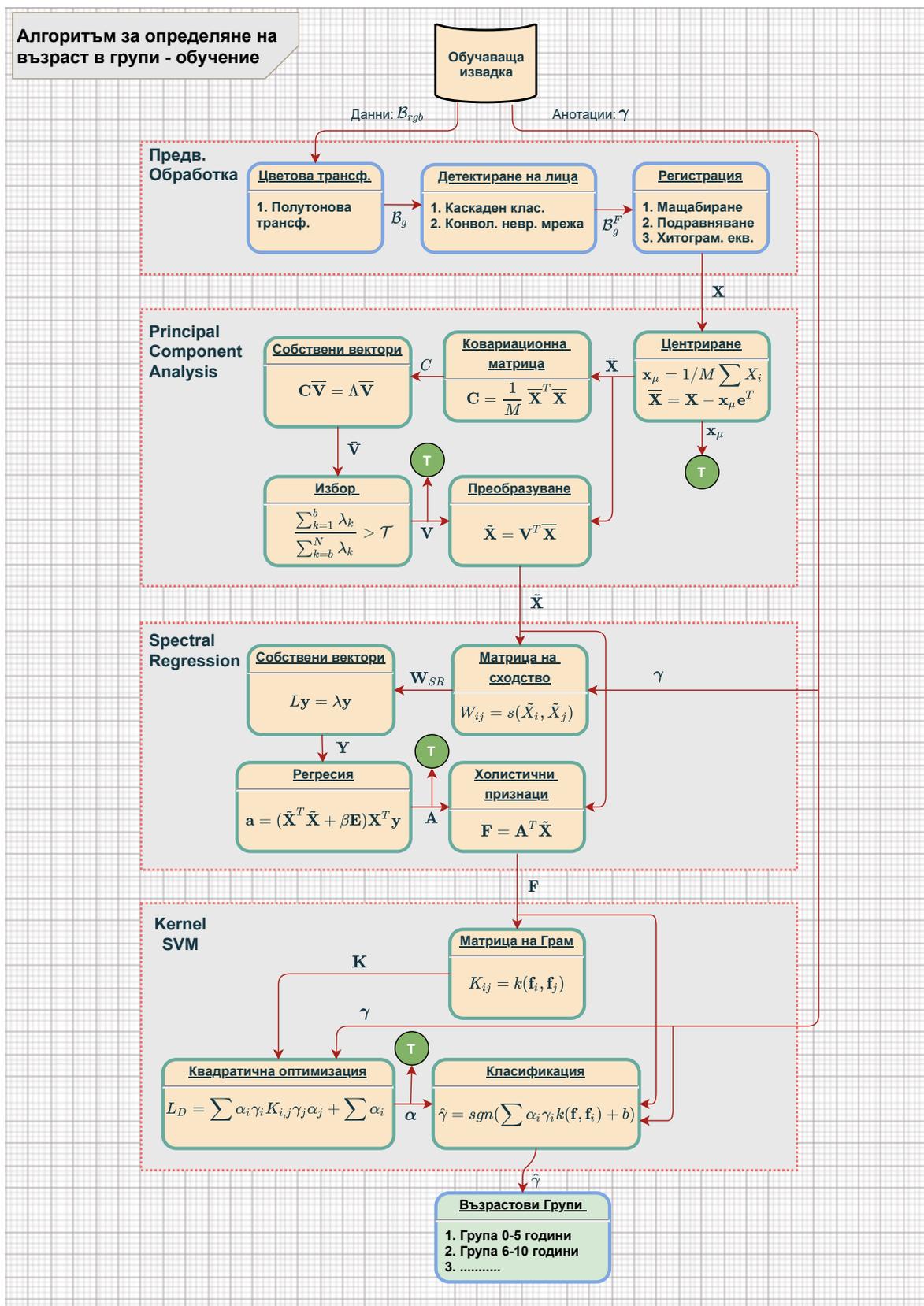
Класификацията на входно изображение в предварително зададени възрастови групи преминава през няколко етапа на обработка, и следва общата рамка, описана в т. 1.1.2. Методите които се използват за да се реализира класификацията се базират на обучение, от което следва, че коефициентите на операторите на всеки един от тях трябва да се изчислят с обучаваща извадка. Поради това, алгоритъмът се реализира в две фази, които са Обучение и Тест. Блок-схема на потока на информацията във фазата обучение е представен на Фиг. 2.1.

При обучението се използва обучаваща извадка, която представлява изображения на лица на естествен фон, всяко от което е ръчно аотирано с възрастова група на индивида. Аотирането се извършва в момента на заснемане на изображението, като се записва възрастта на индивида. Така, обучаващата извадка съдържа изображения на лица \mathcal{B}_{rgb} , който брой може условно да се приеме за M , и аотации γ , които представляват възрастта на индивидите. Изображенията на лицата първоначално се предават на етапа **Предварителна Обработка** (Фиг. 2.1) в който се прави подготовка, състояща се в детектиране и регистриране на лицата. Тъй като цветът на кожата на лицето не представлява фактор при определянето на възраст, изображенията се трансформират в полутонови. Тази трансформация е важна и за детектора на лица, защото той използва само структурата на лицето, но не и неговата цвятова информация. Полутоновото преобразуване се прилага поелементно (за всеки пиксел) за всяко изображение от \mathcal{B}_{rgb} чрез следната формула:

$$\mathbf{V}_G(m, n) = 0.2989 * \mathbf{V}_r(m, n) + 0.5870 * \mathbf{V}_g(m, n) + 0.1140 * \mathbf{V}_b(m, n), \quad (2.1)$$

където \mathbf{V}_r , \mathbf{V}_g и \mathbf{V}_b са съответно червената, зелената и синята цвятови компоненти на изображението, а (m, n) са пространствените координати на пиксела. След полутоновото преобразуване се прилага алгоритъмът за детектиране на лица. Детектирането на лица е описано в т. 2.2.2, при което се цели отделяне на лицето от фона. Детектираните лица се представят чрез координатите на правоъгълника, заграждащ лицето, който се подава на блока *Регистрация*. В този етап, първоначално се изрязва областта на лицето от цялото изображение, използвайки заграждащият правоъгълник на лицето. След това, лицето се мащабира и подравнява така, че очните базиси на всяко лице от обучаващата извадка да съвпадат. Размерите на изображенията им също се уеднаквяват, като броят на пикселите им условно може да бъде приет за N . С тази предварителна обработка се адресира необходимостта от детектиране и изрязване на лицето, за да се отстрани фона, както и регистрацията му за да се намалят вариациите от позата. Необходимостта от тези обработки беше установена от анализа на литературата извършен в т. 2.1.

Регистрираните изображения на лицата се представят като вектор-стълб и се нареждат в матрицата \mathbf{X} , която е с размери $N \times M$. Тази матрица се подава на етапа **Principal Component Analysis** с основна цел намаляване на размерността и отстраняване на шум (виж. т. 1.2). Първата стъпка от този алгоритъм е центрирането на данните, в която се изчислява средната стойност на обучаващата извадка $\bar{\mathbf{x}}_\mu$ и се изважда от тях. Използвайки центрираната матрица $\bar{\mathbf{X}}$ се



ФИГУРА 2.1: Блок-схема на предложенят алгоритъм за класификация на изображения на лица във възрастови групи - фаза Обучение.

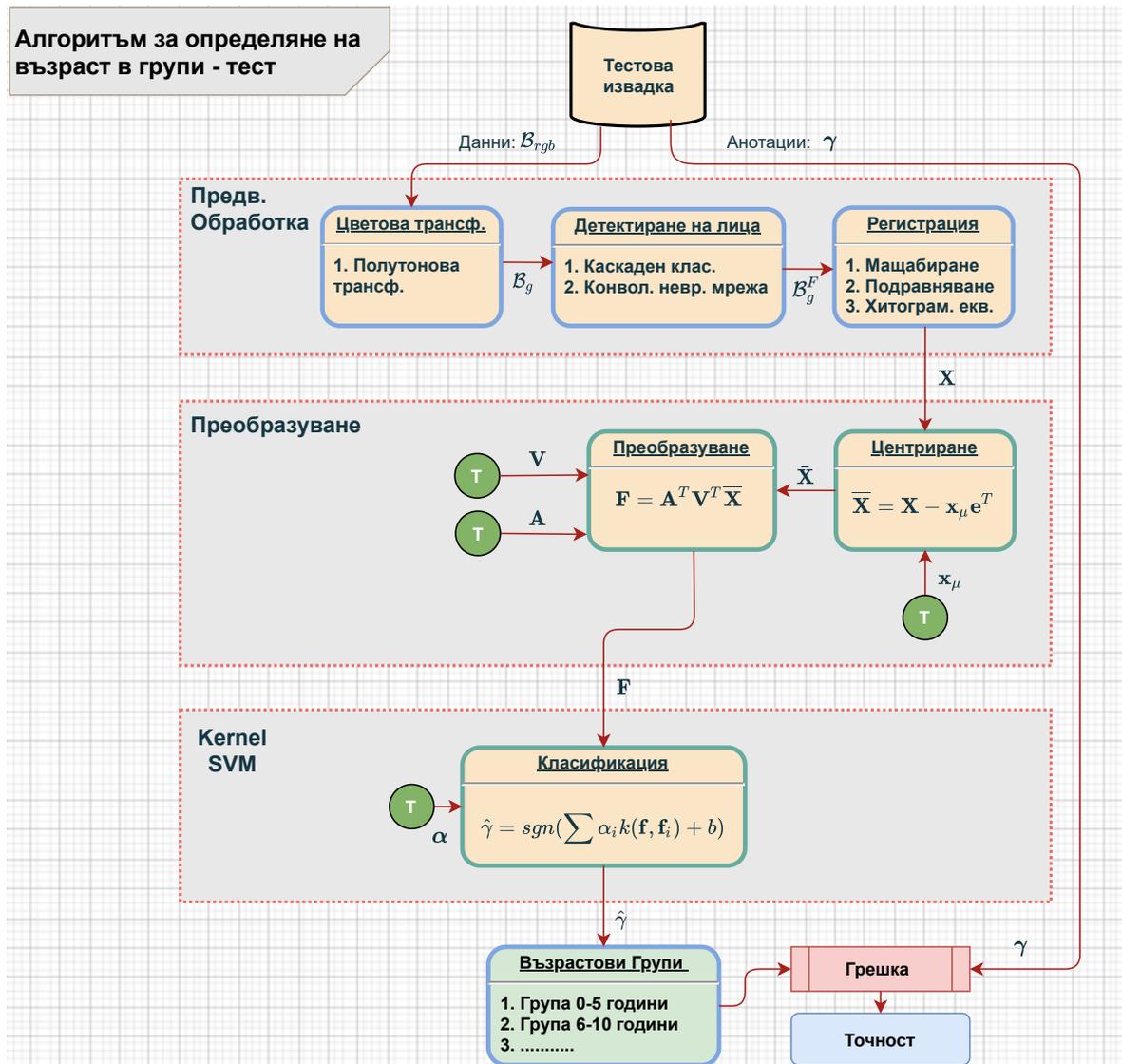
изчислява ковариационната матрица \mathbf{C} , на която се изчисляват собствените вектори $\bar{\mathbf{V}}$. Водещите собствени вектори се избират в стъпката *Избор*, чрез правило и праг \mathcal{T} , и с тях се построява матрицата на преобразуването \mathbf{V} . Именно с тази матрица, в стъпката *Преобразуване*, се преобразуват данните, като се проектират в подпространство с по-ниска размерност, и по този начин се постига намаляване на размерността. Средната стойност на обучаващата извадка \mathbf{x}_μ , и базисът на преобразуването \mathbf{V} , се запазват за фазата Тест, което е обозначено на фигурата със зелен кръг, в центъра на който има буквата Т. Целта на прилагането на PCA е да се намали размерността за да се избегне „проклятието на размерността”, защото изображенията на лицата са с много висока размерност, равна на броя на пикселите. Мотивацията за прилагането на PCA се базира на изводите от предходната точка от тази глава, както и анализа на литературата направен в Глава 1.

Преобразуваните от PCA данни $\tilde{\mathbf{X}}$ се подават на етапа **Spectral Regression**, в който се реализира алгоритъмът Spectral Regression (SR). Този алгоритъм, е от класа които генерират признакови пространства от тип многообразие, и с който се извличат холистични признаци. Това се постига, като многообразието определено от $\tilde{\mathbf{X}}$ се вгражда в евклидово пространство с по-малка размерност, запазвайки структурата му. При първата стъпка от SR се изчислява матрица на сходство \mathbf{W} , използвайки разстоянията между данните и анотациите за възрастовите групи. Тази матрица след това се разлага на собствени вектори \mathbf{Y} , като се решава задачата за собствените стойности.

Използвайки водещите собствени вектори, в стъпката *Регресия*, се изчислява базиса на преобразуването \mathbf{A} чрез квадратична регресия. Този базис, в стъпката *Холистични признаци*, се използва за да се извлекат холистичните признаци \mathbf{F} , чрез проектиране в подпространство. Сама по себе си тази стъпка е също свързана с намаляване на размерността, като размерността на \mathbf{F} е $M \times F$, а $F < N$. Базисът \mathbf{A} се запазва за фазата Тест, както е обозначено на фигурата. Мотивацията за извличане на холистични признаци чрез SR, в предложеният алгоритъм, е следната:

1. Моделирането на възрастта чрез многообразия е доказан метод, както се установи от анализа на литературата в тази глава.
2. Методът позволява да се интегрира дискриминативна информация, което доказано подобрява ефективността на алгоритмите на база на изводите, направени в Глава 1.
3. Методът използва регуляризация, като част от линеаризацията, което въвежда устойчивост на шум, използване на разредени решения, и устойчивост на решенията при малка извадка.
4. Метода с ядрото лесно се интегрира, като част от алгоритъма, която позволява постигане на нелинейност.

Последният етап от предложеният алгоритъм, **Kernel SVM**, се използва за да се класифицират холистичните признаци \mathbf{F} във възрастовите групи с алгоритъма Support Vector Machine, реализиран по метода с ядрото. Във фазата на обучение на този алгоритъм се изчисляват т.нар. опорни вектори и техните коефициенти,



ФИГУРА 2.2: Блок-схема на алгоритъм за класификация на изображения на лица във възрастови групи - фаза Тест.

обозначени на схемата с α . Опорните вектори се изчисляват чрез квадратична оптимизационна задача в стъпката *Квадратична оптимизация*, използвайки предварително изчислената матрица на Грам \mathbf{K} , която от своя страна се изчислява като се използва ядрото k . Изчислените опорни вектори се запазват за фазата Тест, както е обозначено на фигурата. Мотивацията за използването на SVM се дължи на това, че той се счита за оптимален класификатор, има надеждни имплементации с отворен код, и позволява използване на метода с ядрото [105].

Блок-схема, на алгоритъма за определяне на възраст в групи, на фазата Тест, е представена на Фиг. 2.2. Изчислените във фазата обучение коефициенти на PCA преобразуването \mathbf{V} , средната стойност \mathbf{x}_μ , коефициентите на SR преобразуването \mathbf{A} и опорните вектори на SVM α , се използват във фазата Тест, за да се оцени ефективността на алгоритъма. Тестовата извадка по принцип е количествено по-малка от обучаващата извадка и тя също съдържа анотациите по възрастови групи. Първоначално се извършват всички предварителни обработки по абсолютно

същия начин, както във фазата Обучение в етапа **Предварителна Обработка**. След това се прилага центриране и преобразуване на данните в етапа **Преобразуване**. Преобразуването се състои в умножение на центрираните данни \bar{X} с V и s , с което се извличат холистичните признаци F . Използвайки опорните вектори α от фазата на обучение, се прави определяне на възрастовата група $\hat{\gamma}$, въз основа на холистичните признаци. Определената възрастова група се сравнява с истинната такава, съхранявана във вектора γ , което сравнение задава и грешката на алгоритъма, блока *Грешка*. При вярно определена възрастова група, грешката е 1, докато при грешно определена, грешката е 0. Грешката акумулирана за всички индивиди (всяка итерация от LOPO), и техните изображения и разделена на броя на тестовите изображения за всички индивид, дава точността, което е и ефективността на алгоритъма (блок *Точност*).

Подробно описание на всяка една стъпка от алгоритъма за класификация във възрастови групи е представена в следващите подточки.

2.2.2 Детектиране на лица в изображения на естествени сцени с комбиниран класификатор

Алгоритъма за детекция на лица се базира известния алгоритъм за детекция на лица [106], наименуван Viola-Jones на авторите му. Предложеният в тази точка алгоритъм е реализиран с каскада от слаби класификатори и последователно включена конволюционна невронна мрежа Convolutional Neural Network (CNN). Използването на каскада от слаби класификатори позволява реализирането на високо бързодействие на алгоритъма, породено от бързото отхвърляне на обекти принадлежащи на фона и концентрирането му върху обекти, сходни на човешки лица. Това бързодействие е за сметка на точността на детектора, поради което се налага последваща обработка за подобряване на качеството му. Този тип детектори често се комбинират с по-точни класификатори в комбинирана каскада от класификатори от две нива, за да се отстранят характерния за каскадните класификатори голям брой фалшиви сработвания, тоест регистриране на фон като лице.

Каскадният класификатор приема на входа си полутновото изображение B_g . Детекцията на лицата се извършва като се обхожда това изображение с плъзгащи се прозорци $\mathbf{o}_{m,n}$ с различен мащаб, с което се постига детектиране на лица с различни размери, където (m, n) са координатите на пикселите в изображението. В зависимост от настройките, тези прозорци могат да се припокриват както и да са ограничени по отношение на минималния и максималния размер който могат да приемат. Припокриването води до по-голяма точност на детектора, но за сметка на занижено бързодействие. От всеки прозорец се извличат признаци като се използва уейвлетното преобразуване на Нааг [107]. Характерното за този тип уейвлети е, че те могат да се реализират само с операциите събиране и изваждане което значително ускорява изчисленията. Иновативно решение което реализира ускорено преобразуването на Нааг наречено интегрално изображение е предложено в [106]. То се изчислява предварително и позволява да се ускори значително обхождането на изображението както и реализирането на различни скали на уейвлетите.

Върху извлечени признаците от всеки прозорец се прилага комбинацията от каскадните класификатори за да се вземе решение дали съответният признак принадлежи на лице или не. Тази комбинация се изразява чрез:

$$H(\mathbf{o}_{m,n}) = \frac{\sum_{p=1}^P \alpha_p h_p(\mathbf{o}_{m,n})}{\sum_{p=1}^P \alpha_p}, \quad (2.2)$$

където $\alpha_p \geq 0$ са коефициентите на изпъкналата (осигурява се от знаменателя) комбинация. Функциите h_p са т.нар. слаби класификатори с които се взема решение сравнявайки p -тия признак с праг:

$$h_p(\mathbf{o}_{m,n}) = \begin{cases} +1 & \text{ако } u_p(\mathbf{o}_{m,n}) > \tau_p; \\ -1 & \text{в другият случай,} \end{cases} \quad (2.3)$$

където $u_p(\mathbf{o}_{m,n})$ е p -тия признак, а τ_p е асоциирания p -ти праг който се изчислява по време на обучение. Функцията за вземане на решение, дали прозорецът който се анализира е лице или не, е базирана на уравнението (2.2) и е следната:

$$\hat{\eta} = \text{sign}(H(\mathbf{o}_{m,n})), \quad (2.4)$$

т.е. при стойности на $\hat{\eta} > 0$ входния прозорец $\mathbf{o}_{m,n}$ се класифицира като лице и обратно. Тоест, ако знака на изпъкналата комбинация на решенията на всички каскадни класификатори е по-голяма от 0 то прозореца се класифицира като лице и обратно.

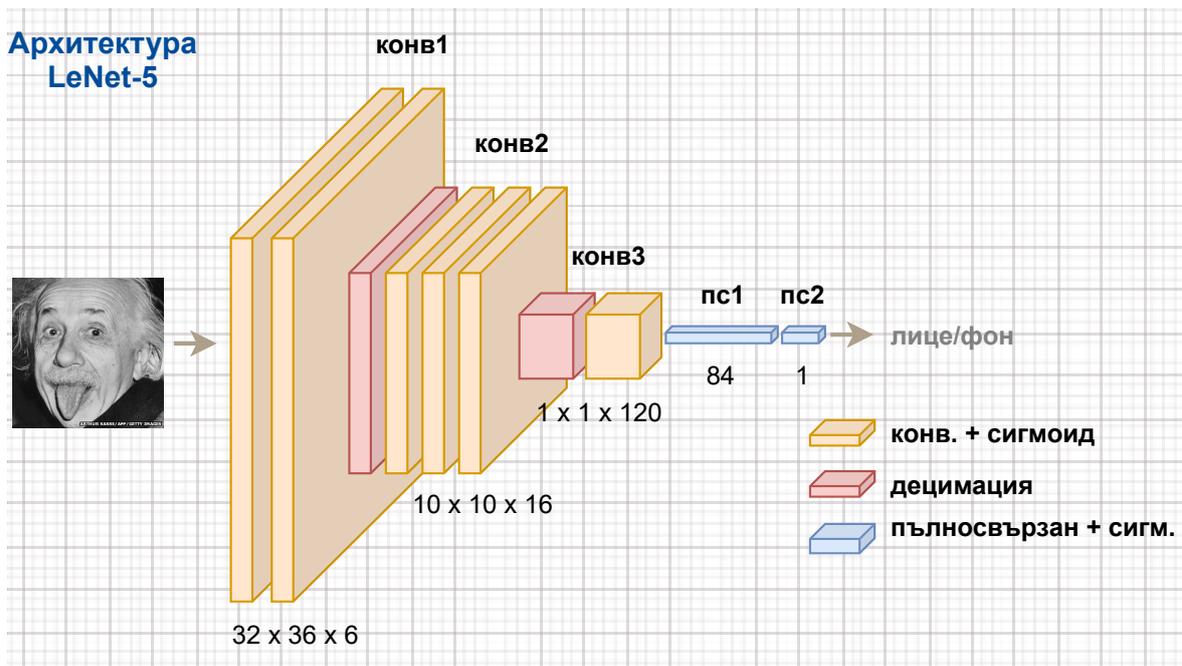
Обучението на този тип класификатори се извършва чрез процедурата Adaboost [108] като основната стъпка от обучението е определянето на коефициентите α_p и на праговете за вземане на решение на слабите класификатори h_p . Тази процедура постига ефективно обединение на слабите класификатори и коефициентите посредством минимизиране на грешката от класификация на H [109]:

$$\arg \min_{\alpha_p, \tau} J(H) = \sum_q \exp^{-\eta_q H(\mathbf{o}_{m,n})} = \sum_q \exp^{-\eta_q \sum_{p=1}^P \alpha_p h_p(\mathbf{o}_{m,n})}, \quad (2.5)$$

където $\{\mathbf{o}_q, \eta_q\}$ са q -тия прозорец със съответния му етикет $\eta_q \in \{-1; 1\}$ от обучаващата извадка. Грешката на класификация на всяко едно изображение е претеглена експоненциално с което се апроксимира идеалната $1/0$ функция на загубите [109].

Следващата стъпка от предложеният алгоритъм е вторична класификация на прозореца който се анализира. Ако $\mathbf{o}_{m,n}$ бъде класифициран като фон, то обхождането на изображението продължава, но ако се класифицира като лице, прозореца се подава към класификатора реализиран с CNN. Основен принос в разработката на алгоритъма за детектиране на лица, е добавянето на втори етап на класификацията реализиран с CNN. Структурата на невронната мрежа е с разреждана, а не с пълна свързаност и следва предложената в [110] архитектура наименувана LeNet-5. Така са редуцирани слоевете, извършващи едновременно конволюция и децимация във всяко рецепторно поле, с което се повишава общата скорост на обработка с конволюционната невронна мрежа.

CNN взема решение въз основа на индивидуалните изходи от невроните като,



Фигура 2.3: Архитектура на конволюционната невронна мрежа LeNet-5 [110].

за вход приема изображението на лицето след като е класифицирано като такова от каскадния класификатор. Претеглената сума от входовете на неврона за слой l и област p се изчислява с:

$$S^{l,p}(\mathbf{o}_{m,n}) = \sum_{q=1}^Q \zeta^{l-1,q}(\mathbf{o}_{m,n}) * \omega^{l,p,q} - \mathbf{b}^{l,p} \quad (2.6)$$

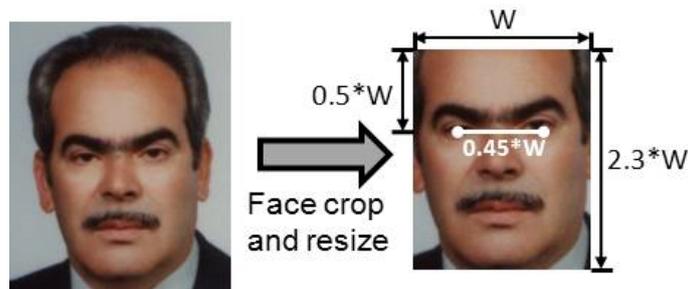
където $\omega^{l,p,q}$ са теглата на неврона за слой l , област p и карта q , $\mathbf{b}^{l,p}$ е прагът на невроните за от карта p и слой l а Q е броят на картите. Изходът от неврон за слой l и област p се определя от:

$$\zeta^{l,q}(\mathbf{o}_{m,n}) = \Phi(S^{l,p}(\mathbf{o}_{m,n})) \quad (2.7)$$

където Φ е биполярната сигмоидална активиращата функция с изход в интервала $[-1; 1]$ [110]. Изходът от последния неврон на мрежата определя дали подадения входен прозорец е лице или фон в зависимост от, това дали е положителен или отрицателен. Невронната мрежа се състои от 7 слоя и на входа се подава полутоново изображение с размери 32×32 (изображението се преоразмерява преди да се подаде). Архитектурата на мрежата е представена на Фиг. 2.3.

Мрежата се състои от три конволюционни слоя, конв1, конв2 и конв3 с размерности съответно $32 \times 36 \times 6$, $10 \times 10 \times 16$, $1 \times 1 \times 120$ и два пълносвързани с размерности съответно 84 и 1. Между пълносвързаните слоеве има децимация реализирана с маска 2×2 . На изхода на всеки конволюционен слой се прилага сигмоидната функция.

Обучението е зададено в два класа, лица и фон. Допълнително, изображенията на лица, с които е обучена CNN варират в интервалите $[-90, +90]$ ротация и $[-45,$



ФИГУРА 2.4: Визуализация на процеса по подравняване на лица.

45] наклон, което предоставя възможност за работа с широк диапазон от сцени. Обучението на невронната мрежа се извършва чрез алгоритъма за обратно разпространение на грешката, като всички тегла в нея се преизчисляват едновременно във всяка итерация [110]. Това позволява обучение на мрежата в експлоатационния режим в случай, че не всички данни от обучението са налични от самото начало.

2.2.3 Регистрация на изображения на лица

В повечето системи за анализ на човешки лица, регистрацията се счита за критична стъпка от предварителната обработка. Тя може да включва геометрично подравняване, нормализация на осветеността или в някои случаи отстраняване на шум. В предложеният алгоритъм се прилага геометрично подравняване и нормализация на осветеността. Първото намалява вариациите по отношение на позата, а при второто се намаляват вариациите, които се генерират от различните осветености. Чрез регистрацията се цели запазване само на вариациите които са характерни при оценката на възрастта, като например геометрична форма на лицето и бръчки.

Геометричното подравняване, в предложеният алгоритъм, се реализира с очният базис. За формирането на този базис, първоначално се детектират центровете на очите, използвайки алгоритъма, предложен в [111]. За да се увеличи точността на детектиране на окото, търсенето на лявото и дясното око са ограничени съответно в горната лява и дясна половици на изображението. След като са открити центровете на двете очи формално се задава линията която ги свързва и която определя очният базис. Тази линия се използва за изчисляване на ъгъла на наклон на лицето спрямо хоризонталната ос на изображението. Ако този ъгъл е по-голям от ± 25 градуса, тогава лицето се отхвърля и не се използва. Ако този ъгъл е в порядъка от ± 5 до ± 25 градуса, тогава лицето се счита за наклонено и използвайки изчисленият ъгъл се прилага ротация на изображението чрез афинна трансформация за да се хоризонтира. Ако ъгълът е в порядъка на ± 5 градуса, тогава лицето се счита за ненаклонено и не се прилага ротация. След като се завърти лицето така, че очният базис да е в границите ± 5 градуса спрямо хоризонталната ос, се преминава към позициониране и мащабиране. Това става с помощта на скалиране спрямо широчината на изображението в пиксели W , която лицето трябва да приема и която е зададена предварително. Процеса е визуализиран на Фиг. 2.4.

Критерият за подравняване на лицата се изпълнява като първо се оразмери лицето така, че очният базис да е с размер $0.45W$ в пиксели. След което лицето се

транслира така, че центъра на очният базис да минава през линията която дели изображението на две по вертикала, тоест $0.5W$, което осигурява подравняване по хоризонтала. Аналогично, лицето се транслира по вертикала така, че очният базис да е на разстояние $0.5W$ спрямо горният ляв ръб на изображението, тоест началото на координатната му система. След като бъде оразмерено и подравнено лицето се изрязва така, че големината му по хоризонтала да е W , а по вертикала $2.3W$ в пиксели. Коефициентите с които се извършва мащабирането и подравняването са избрани експериментално, използвайки база данни от лица.

2.2.4 Намаляване на размерността с Principal Component Analysis

Алгоритъмът PCA е от типа оператори които са с обучение, тоест използват статистическите зависимости на данните за да изчислят коефициентите на оператора. Поради това, методът се реализира в две фази: обучение и тест. Обучението се извършва с обучаваща извадка, а теста - с тестова извадка. Описанието на алгоритъма и математическите означения, представени на Фиг. 2.1, съвпадат с описанието в тази точка и фигурата може да се използва за референция.

Обучаващата извадка е представена като матрица $\mathbf{X} \in \mathbb{R}^{N \times M}$, където M е броят на изображенията, а N е броят на пикселите в изображенията. Първата стъпка от алгоритъма е центриране на обучаващата извадка със средната стойност на данните, която се изчислява чрез:

$$\mathbf{x}_\mu = \frac{1}{M} \sum_{i=1}^M X_i \quad (2.8)$$

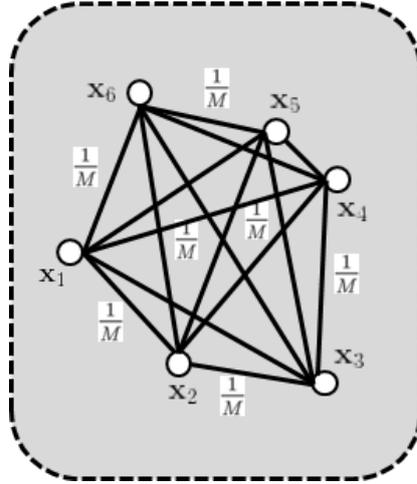
Това практически означава, намиране на средна стойност на колоните на матрицата \mathbf{X} . Средната стойност представлява намиране на емпиричната средна стойност на гаусовото разпределение от което са извлечени данните. Центрирането се извършва като от всяка колона от матрицата на данните се извади средната стойност:

$$\bar{\mathbf{X}} = \mathbf{X} - \mathbf{x}_\mu \mathbf{e}^T, \quad (2.9)$$

където $\mathbf{e} \in \mathbb{R}^M$ е единичният вектор. С центрирането се цели подготовка за работа с ковариационната матрица на гаусовото разпределение, което е и следващата стъпка от алгоритъма, а именно изчисляване на ковариационната матрица чрез следната формула:

$$\mathbf{C} = \frac{1}{M} \bar{\mathbf{X}} \bar{\mathbf{X}}^T = \frac{1}{M} \mathbf{X} (\mathbf{I} - \frac{1}{M} \mathbf{e} \mathbf{e}^T) \mathbf{X}^T \quad (2.10)$$

където \mathbf{I} е единичната матрица. Първия запис е стандартното приближение на ковариационната матрица с краен брой данни, равен на M . Вторият запис позволява интерпретация чрез вграждането на графи, въведено в т. 1.2.6, където изразът в скобите представлява матрицата на Лаплас \mathbf{L}_{PCA} със съответната тегловна матрица $\mathbf{W}_{PCA} = \frac{1}{M} \mathbf{e} \mathbf{e}^T$, съответстваща на пълно свързан граф с тегла по ръбовете $\frac{1}{M}$, илюстриран на Фиг. 2.5. Пълносвързаният граф означава, че PCA алгоритъма възприема многообразието на входните данни по такъв начин, че всеки елемент



ФИГУРА 2.5: Примерна диаграма на пълноsvързаният граф определящ PCA съставен от 6 възела.

е с еднакво значение спрямо всички останали. Основната част от алгоритъма се състои в намирането на направленията в пространството на изображенията които са най-информативни. От статистическа гледна точка това означава намиране на направления, при които данните са максимално не-корелирани и за които, вариацията от проекциите на данните е най-голяма [112]. Това позволява да се използва рамката за моделиране с графи и задачата да се дефинира като се търсят направленията с минимална вариация, които са ортогонални и с норма единица [79]:

$$\bar{\mathbf{v}} = \arg \min_{\mathbf{v}^T \mathbf{v} = 1} \mathbf{v}^T \mathbf{C} \mathbf{v} = \arg \min_{\mathbf{v}^T \mathbf{v} = 1} (\mathbf{X} \mathbf{v})^T \mathbf{L}_{PCA} (\mathbf{X} \mathbf{v}) \quad (2.11)$$

Очевидно е, че PCA следва линеаризацията на вграждането на графи, използвайки основния граф, свързващ всички двойки данни с равни тежести и с ограничения върху нормата на проекционния вектор. Решението на (2.11) се намира чрез решаването на задачата за собствени стойности:

$$\mathbf{C} \bar{\mathbf{V}} = \Lambda \bar{\mathbf{V}}, \quad (2.12)$$

където $\bar{\mathbf{V}}$ е матрицата със собствените вектори, а Λ е диагонална матрица съдържаща съответните собствени стойности, подредени по намаляващ ред. Тъй като ковариационната матрица е положително определена, собствените стойности са положителни и всяка една отговаря на дисперсията (квадрата от вариацията) по направлението, определено от съответния ѝ собствен вектор. Следователно, по-голяма собствена стойност отговаря на по-информативно направление.

След като са изчислени собствените вектори и стойности, следващата стъпка от алгоритъма PCA е избор на най-информативните направления. Това става чрез избор на водещите собствени вектори, представляващи най-информативните направления използвайки следния критерий:

$$RMSE(d) = \frac{\sum_{k=1}^d \lambda_k}{\sum_{k=1}^N \lambda_k} > \mathcal{T} \quad (2.13)$$

където \mathcal{T} е прага с който се означава желаното отношение. Този критерий използва факта, че собствените стойности са положителни числа, и образува отношението на сумата на избраните собствени стойности към сумата на всички собствени стойности. Това означава, че в зависимост от избраният брой d , Root Mean Square Error (RMSE) $RMSE(d)$ варира от 0 до 1, има интуитивна интерпретация и позволява да се работи в проценти. Обикновено прага \mathcal{T} се избира така, че това отношение да бъде в границите между 0.85 и 0.95. Това означава, че направления с малки вариации се премахват което се интерпретира като премахване на шум или вариации в изображенията които не са съществени за задачата. При прилагането на този критерий се предполага, че собствените стойности са наредени в низходящ ред спрямо индекса k . Параметърът d се избира така, че да е минималният брой собствени стойности удовлетворяващи неравенството (2.13). Именно този брой задава размерността на пространството с намалената размерност.

След като бъдат избрани собствените вектори, отговарящи на най-информативните направления, те се подреждат в матрицата $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d]$. Тази матрица представлява базиса, с който се извършва проекция в подпространство, и се използва както във фазата Обучение за да се подготвят данните за следващите етапи от обработката, така и във фазата Тест. Преобразуването използва центрираните данни и то се извършва по следният начин:

$$\tilde{\mathbf{X}} = \mathbf{V}^T \bar{\mathbf{X}} \quad (2.14)$$

Като резултат от тази операция матрицата с данните, която се подава на следващата стъпка е с размерност $\tilde{\mathbf{X}} \in \mathbb{R}^{d \times M}$.

2.2.5 Извличане на холистични признаци с алгоритъма Spectral Regression

След като данните се проектират в пространство с намалена размерност от алгоритъма PCA, е необходимо да се осигури тяхната дискриминативност (PCA алгоритъмът отстранява шум което не гарантира дискриминативност). Това означава, те да се подготвят за процеса на класификация, който да бъде реализиран ефективно и ефикасно. Дискриминативност на данните означава малка вътрешнокласова вариация и голяма между класова вариация. Ако данните са с ниска дискриминативност, то тогава е необходимо да се използва класификатор, който може да моделира по-сложни разделящи хиперповърхнини, което е за сметка на по-голяма изчислителна сложност и обратно, ако данните са дискриминативни то класификатора може да бъде по-прост. Целта на използването на алгоритъма SR за определяне на възраст в групи е именно да се увеличи тяхната дискриминативност, като също така се намали размерността. Този алгоритъм апроксимира многообразието от което са извлечени данните чрез граф, след което намалява размерността им, запазвайки техните геометрични зависимости. Нещо повече, използвайки информацията за анотациите по възрастовите групи при построяването на графа се гарантира увеличена дискриминативност в пространството с намалена размерност. Описанията и означенията, за етапа **Spectral Regression** на Фиг. 2.1, съвпадат с тези в тази точка.

Предложен от [113] алгоритъмът SR служи за извличане на холистични признаци и поради намаляването на размерността на изходните данни, той може да се разглежда и като алгоритъм за намаляване на размерността. Характерно за него е, че се преформулира решаването на задача за собствени стойности от вграждането на графи, в задача за регресия, което му позволява по гъвкаво построяване на матрицата на преобразуването използвайки регуляризация. По-точно, основните му предимства са следните:

- Решаването на задача за регресия е изчислително по-бързо, което е предимство при работа с големи обеми от данни.
- Регресията може да бъде решена с въвеждането на регуляризационен член, който: контролира сложността на решението; спомага за избягването на претрениране; може да осигури разреденост на решението.
- Подходящ е при малък брой на данните от извадката, но голяма размерност на изображенията.
- Запазва характерните особености на линейните статистически преобразувания и тяхната реализация с ядро.

Второто предимство е това, което отличава SR като самостоятелен алгоритъм, който въвежда допълнителен елемент при анализа на данни, поради което е избран като компонент от алгоритъма за класификация във възрастови групи. Регресията може да бъде формулираната чрез различни целеви функции, както и контрол на сложността на търсените на проектиращи направления чрез регуляризация. За разлика от PCA, където се търсят направления, които максимизират вариацията на проектираните данни, SR търси направления при които се запазва геометричната структура на данните. В общия случай тази геометрична структура е многообразие (виж т. 1.2.5) поради което, признаците извлечени от SR са такива, при които приблизително се запазва геометричната структура в пространството с намалена размерност. Алгоритъмът също така се описва и от рамката представена в т. 1.2.6.

На първата стъпка от алгоритъма се построява тегловната матрица \mathbf{W}_{SR} като се използват данните $\tilde{\mathbf{X}}$ с намалена от PCA алгоритъма размерност, и съответните им етикети γ . Това е симетрична матрица с размери $M \times M$ съдържаща информация за сходството между всеки два вектора от $\tilde{\mathbf{X}}$. За изчисляването на сходството освен етикетите се използва и геометрична близост между всеки два вектора. Така, за построяването на \mathbf{W}_{SR} се използва всяка двойка вектори от $\tilde{\mathbf{X}}$, $\tilde{\mathbf{x}}_i$ и $\tilde{\mathbf{x}}_j$, индексирани с i и j , като всеки един от елементите на \mathbf{W}_{SR} приема една от следните стойности:

- $W_{i,j} = 1/l_k$ ако $\tilde{\mathbf{x}}_i$ и $\tilde{\mathbf{x}}_j$ заедно принадлежат на k -тия клас, където l_k е броя на елементите в него;
- $W_{i,j} = \delta s(i, j) = \delta \exp\left(-\frac{\|\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j\|}{\sigma}\right)$ ако $\tilde{\mathbf{x}}_i$ е сред трите най-близките съседи на $\tilde{\mathbf{x}}_j$ или обратно. Функцията $s(i, j)$ изчислява сходството между $\tilde{\mathbf{x}}_i$ и $\tilde{\mathbf{x}}_j$, за $\sigma \in \mathbb{R}$. Параметъра δ се използва за настройка на важността между елементите които имат етикети и тези, които имат голямо сходство, изчислено чрез $s(i, j)$. Обикновено $\delta = 0.1$, а σ се изчислява експериментално;

- $W_{i,j} = 0$ във всички останали случаи.

Матрицата \mathbf{W}_{SR} задава не-пълно свързан, не-насочен граф, който описва както геометричната структура на данните, така и тяхната класова принадлежност. Класовата принадлежност се задава като се използват етикетите и броят на изображенията в даден клас. По този начин се определят елементите от всеки клас като „свързани”. За да се запази геометричната структура, всеки от елементите се свързва със съседните му три, които са най-близки по евклидово разстояние (използва се алгоритъма k-NN). В този случай обаче теглата се определят от функцията на сходство $s(i, j)$, която е нелинейно намаляваща с увеличаване на разстоянието, а скоростта ѝ на намаляване се регулира от параметъра σ . Ако броят на данните във всеки клас е много голям, то тогава теглата с които се определя класовата принадлежност ще станат много малки спрямо теглата с които се определя геометричната структура на многообразието. За да се балансира се въвежда параметъра δ , с който се регулира „важността” на тези сходства спрямо свързаността, определена от анотациите. В случаите когато векторите не споделят общ клас или не са сред трите най-близки съседа се счита, че между тях няма свързаност и теглата се приравняват на нула.

Използвайки \mathbf{W}_{SR} се построява матрицата graph-Laplacian $\mathbf{L}_{SR} = \mathbf{D} - \mathbf{W}_{SR}$ където $D_{j,j} = \sum_i W_{i,j}$. Следващата стъпка е решаването на задачата:

$$\mathbf{y} = \arg \min_{\mathbf{y}^T \mathbf{D} \mathbf{y} = 1} \mathbf{y}^T \mathbf{L}_{SR} \mathbf{y}, \quad (2.15)$$

където решенията \mathbf{y} представляват вграждането на многообразието и се намират чрез задачата за собствените стойности и вектори. Тази стъпка отличава SR от рамката, приета при намаляването на размерността с вграждане на графи. Основното му предимство е, че размерността на \mathbf{W}_{SR} е равна на броя на данните, която на практика е много по-малка от размерността на изображенията, от където идва и оптимизация по отношение на изчисленията. Важно е да се подчертае, че броят на собствените вектори $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_r]$ се избира да бъде r , равен на броя на класовете минус 1 като това се обосновава на база предишни изследвания [114].

За разлика от PCA, избраните собствени вектори определят т.нар. “soft clusters” в изходното пространство, а не в пространството на данните. За да се възстановят направлениата с които се извличат признаци се прилага следващата важна стъпка от SR, която е именно регресията:

$$\mathbf{a}^k = \arg \min_{\mathbf{a}} \sum_{i=1}^M (\mathbf{a}^T \mathbf{x}_i - y_i^k), \quad (2.16)$$

където y_i^k е i -тият елемент от k -тият вектор \mathbf{y}^k , като $k = [0, \dots, r]$. В случаите в които има вариации в данните, които са породени от шум, и които не са били отстранени от PCA алгоритъмът, решенията на тази задача могат да бъдат силно изместени. Също така, когато броят на данните е по-малък от размерността на PCA пространството, тази задача няма стабилно решение [113]. Поради това, регресията се решава чрез въвеждането на регуляризация на решението:

$$\mathbf{a}^k = \arg \min_{\mathbf{a}} \left(\sum_{i=1}^M (\mathbf{a}^T \mathbf{x}_i - y_i^k)^2 + \beta \|\mathbf{a}\|^2 \right). \quad (2.17)$$

Регуляризацията в случая се изпълнява чрез L_2 нормата, тоест дължината на вектора, и задаване на важността става с коефициента $\beta \in \mathbb{R}$. В зависимост от задачата обаче, тази норма може да бъде L_1 , с което се генерират разреждени решения \mathbf{a} . Решенията на тази задача се подреждат в матрица $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_r]$, която се явява матрицата на преобразуването.

Когато се търси реализация с ядро, използвайки теоремата за представителя, функцията се задава като:

$$f(\mathbf{x}_i) = \sum_{j=1}^M \theta_j K(\mathbf{x}_i, \mathbf{x}_j), \quad (2.18)$$

използвайки избрано ядро K . Задачата за намиране на функцията на вграждане тогава се формулира като:

$$\mathbf{b}^k = \arg \min_{f \in \mathcal{H}} \left(\sum_{i=1}^M (f(\mathbf{x}_i) - y_i^k)^2 + \beta \|f\|_{\mathcal{H}} \right), \quad (2.19)$$

където y_i^k е i -тият елемент от k -тият вектор, $k = [0, \dots, r]$ и \mathbf{y}^k , \mathbf{b}_k е θ_k , а $\|\cdot\|_{\mathcal{H}}$ е нормата в хилбертовото пространство с възпроизвеждащо ядро \mathcal{H} . Матрицата с решенията, която се използва при преобразуването, се формира като $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_r]$.

Преобразуването за линейният случай се изчислява като:

$$\mathbf{F} = \mathbf{A}^T \tilde{\mathbf{X}}, \quad (2.20)$$

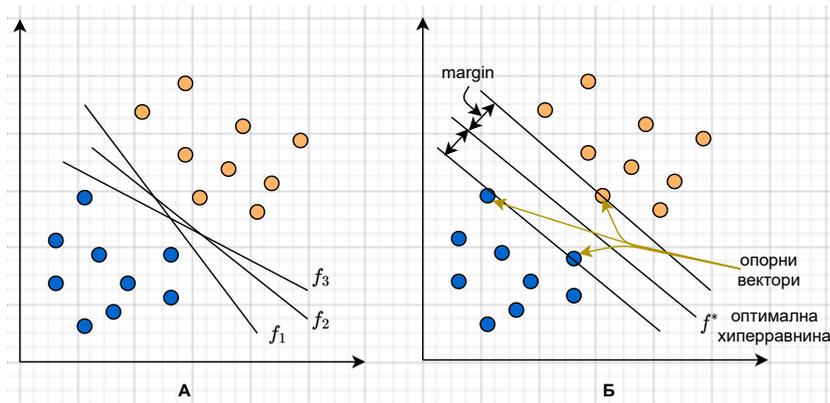
а за варианта с ядро, преобразуването се изчислява като:

$$\mathbf{f}_i = \mathbf{B}^T \mathbf{K}(:, \tilde{\mathbf{x}}_i), \quad (2.21)$$

където \mathbf{f}_i е i -тата колона от матрицата \mathbf{F} , $\tilde{\mathbf{x}}_i$ е i -тата колона от $\tilde{\mathbf{X}}$, а $K(:, \tilde{\mathbf{x}}_i) = [K(\mathbf{x}_1, \tilde{\mathbf{x}}_i), \dots, K(\mathbf{x}_m, \tilde{\mathbf{x}}_i)]^T$. Преобразуваните данни се интерпретират като холистични признаци и се използват в следващата стъпка, която е класификацията във възрастови групи.

2.2.6 Класифициране във възрастови групи с Support Vector Machines

Класификацията във възрастови групи е реализирана с алгоритъма SVM. Алгоритъмът е с обучение и използва холистичните признаци, извлечени с алгоритъма SR, за да намери коефициентите на оператора, с който се извършва класификацията. Работното предположение за SVM класификатора за два класа е, че съществува хиперравнина която разделя пространството на две под-пространства такива, че данните от единият клас се намират изцяло в едното под-пространство, а данните от другият клас в другото. Ако това се постига, то SVM класификаторът намира тази хиперравнина, за която разстоянието от равнината до най-близките



ФИГУРА 2.6: Класификация с хиперравнина в два класа: А) разделящи хиперравнини; Б) оптимална хиперравнина с максимално разстояние.

точки от класовете е максимално. Това е изобразено на Фиг. 2.6. Тези точки са подмножество на данните, които се използват за обучението и се наричат Support Vectors (опорни вектори), от където идва и наименованието на класификатора. Когато двата класа могат да бъдат напълно разделени от хиперравнина, класификацията се нарича Hard-margin. На практика данните рядко могат да бъдат разделени от хиперравнина при което се използва т.нар. Soft-margin класификация, която е най-разпространена в практиката. При нея се допуска наличие на малък брой данни от даден клас да са в половината на другия клас и обратно. Друга особеност на този алгоритъм е, че чрез използването на метода с ядрото, той може да се преобразува в нелинеен класификатор. Това позволява разделяне на класове от произволна повърхнина. Математическите означения в тази точка са същите като тези на Фиг. 2.1 и фигурата може да се използва като референтна.

При обучението на SVM на входа му се подават извлечените холистични признаци \mathbf{F} , както и анотациите на всеки признак γ . Ако допуснем, че класификацията се извършва в два класа, тоест $\gamma_i \in \{-1, 1\}$, то линейният SVM търси хиперравнина разделяща двата класа като се допуска, че съществува такава. Нека функцията на тази хиперравнина е:

$$f(\mathbf{f}_i) = \mathbf{w}^T \mathbf{f}_i + b, \mathbf{f}_i \in \mathbb{R}^r, \mathbf{w} \in \mathbb{R}^r, b \in \mathbb{R} \quad (2.22)$$

където \mathbf{w} е нормалата към тази хиперравнина, а b е отместването спрямо координатното начало. Тогава обучението на линейният SVM от тип Hard-margin се състои в решаването на оптимизационната задача:

$$\begin{aligned} & \underset{\mathbf{w}, b}{\text{minimize}} \quad \frac{1}{2} \|\mathbf{w}\|^2 \\ & \text{subject to: } \gamma_i(\mathbf{w}^T \mathbf{f}_i + b) \geq 1, i = 1, \dots, r, \end{aligned} \quad (2.23)$$

където $y_i(\mathbf{w}^T \mathbf{f}_i + b) \geq 1$ определя достижимото множество и гарантира, че всяка точка от обучаващата извадка принадлежи на една и съща половина на подпространството.

Както беше споменато, при практически задачи много рядко се получава перфектна разделимост на класовете. Често, поради случаен шум, се получава да има данни от единия клас на място на което се очаква елемент от другия клас. В този случай се въвежда допускането за малка грешка при класификация, която грешка се контролира при обучението на SVM. Това се постига чрез въвеждането на т.нар. положителни slack променливи, $\xi_i, i = 1, \dots, r$ в ограниченията на оптимизационната задача [105]:

$$\begin{aligned} \mathbf{w}^T \mathbf{f}_i + b &\geq +1 - \xi_i, \text{ for } \gamma_i = +1 \\ \mathbf{w}^T \mathbf{f}_i + b &\leq -1 + \xi_i, \text{ for } \gamma_i = -1 \\ \xi_i &\geq 0 \forall i. \end{aligned} \quad (2.24)$$

При така формулирани ограниченията, за да бъде допусната грешка, тоест данни от единият клас да бъдат класифицирани като данни от другият клас е необходимо ξ_i да бъде по-голямо от 1. Това обаче се явява проблем и затова се въвежда горна граница $\sum_i \xi_i$, с която се контролира броят грешки. Този контрол се осъществява като към целевата функция се добави $\sum_i \xi_i$ и тя се модифицира до:

$$\begin{aligned} \underset{\mathbf{w}, b}{\text{minimize}} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i \\ \text{subject to:} \quad & \mathbf{w}^T \mathbf{f}_i + b \geq +1 - \xi_i, \text{ for } \gamma_i = +1 \\ & \mathbf{w}^T \mathbf{f}_i + b \leq -1 + \xi_i, \text{ for } \gamma_i = -1 \\ & \xi_i \geq 0 \forall i, \end{aligned} \quad (2.25)$$

където C е параметър с който се контролира броят грешки, като при големи C се допускат голям брой грешки, защото се дава по-голяма тежест на този член.

Оптимизационната задача е изпъкнала, и е дори задача от тип квадратично програмиране, което позволява да се използва дуалната формулировка на задачата, тоест очаква се екстремумите на основната и дуална формулировки да съвпадат. Причината да се премине към дуалната формулировка е, че данните от обучаващата извадка ще присъстват само като скаларни произведения между вектори. Това е важно, защото ще позволи използването на метода с ядрото, с което ще се реализира нелинеен SVM. Първата стъпка към дуалната формулировка е построяването на функцията на Лагранж:

$$L_P = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i - \sum_i \alpha_i \{ \xi_i (\mathbf{w}^T \mathbf{f}_i + b) \geq -1 + \xi_i \} + \sum_i \mu_i \xi_i, \quad (2.26)$$

където $\alpha_i, \mu_i, i = 1, \dots, r$ са множителите на Лагранж. След диференциране и прилагане на условията на Karush-Kuhn-Tacker (ККТ) дуалната оптимизационна задача се задава като:

$$\begin{aligned} \underset{\alpha}{\text{maximize}} \quad & \sum_{i=1}^r \alpha_i + \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r \gamma_i \alpha_i (\mathbf{f}_i^T \mathbf{f}_j) \gamma_j \alpha_j \\ \text{subject to:} \quad & \sum_i \alpha_i \gamma_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, \dots, r \end{aligned} \quad (2.27)$$

Решението на така формулирана задачата се дава от N_S на брой не-нулеви α_i които всъщност представляват опорните вектори. Този брой N_S е по-малък от броя на данните r , защото ограниченията в задачата (2.27) задават повечето коефициенти да бъдат нулеви. Това е характерна особеност на алгоритъма която на практика означава, че във фазата Тест, класификацията се извършва като се използват много по-малко данни от тези, които се използват във фазата на обучение, а от там и по-малко изчисления. Решението на основния проблем, използвайки решението на дуалния се представя като:

$$\mathbf{w} = \sum_{i=1}^{N_S} \alpha_i \gamma_i \mathbf{f}_i. \quad (2.28)$$

Във фазата Тест, класификацията за вектора от холистични признаци \mathbf{f} се извършва като се вземе знака на следната функция:

$$\hat{\gamma} = \text{sgn}(\mathbf{w}^T \mathbf{f} + b), \quad (2.29)$$

с което се определя класа.

В дуалната формулировка на задачата (2.27), може да се забележи, че данните присъстват в целевата функция като скаларни произведения. Това позволява да се използва метода с ядрото, при което скаларните произведения $(\mathbf{f}_i^T \mathbf{f}_i)$ се заменят с матрицата на Грам $k(\mathbf{f}_i, \mathbf{f}_i)$ при избрано ядро k . Така класификацията за \mathbf{f} се извършва като:

$$\hat{\gamma} = \text{sgn}\left(\sum_{i=1}^{N_S} \alpha_i \gamma_i k(\mathbf{f}, \mathbf{f}_i) + b\right). \quad (2.30)$$

Тук е важно да се отбележи, че функцията за класификация може да се интерпретира като се използва теоремата за представителя.

2.3 Основни резултати и изводи

Основните резултати от работата в тази глава са следните:

1. Разработен е алгоритъм за детектиране на лица в изображения на естествени сцени чрез комбинация от:
 - Бърз класификатор, реализиран с каскадни класификатори и обучен по метода AdaBoost.
 - Бавен, но точен класификатор, реализиран с конволюционна невронна мрежа, реализирана по тип кодер архитектура и обучена с разширена извадка от лица, включваща големи вариации на позата.
2. Разработен е алгоритъм за определяне на възраст в групи на база изображения на лица, включващ:
 - Регистрация на лицата в обща координатна система, използвайки очният базис.

- Комбинирано намаляване на размерността на изображенията с PCA и извличане на холистични признаци с SR за повишаване на дискриминативността и запазване на геометрията на многообразието, позволяващи реализация по метода с ядрото.

- Класификатор SVM, реализиран с ядро.

Изводите от работата по тази глава са следните:

1. Регистрацията на лицата в обща координатна система са от съществено значение за определянето на възраст;
2. Намаляването на размерността с PCA е от голяма важност, за да се намали шума в изображенията и най-вече да се избегне „проклятието на размерността”.
3. Моделирането на процеса на стареене чрез многообразия е валиден подход с който се постига по-ниска средна грешка при определянето на възрастта.
4. Включването на дискриминативна информация при моделирането на многообразието е от съществено значение при класификацията.

2.4 Приноси към Глава 2

Научно-приложните приноси са:

1. Разработен е подобряващ дискриминативността алгоритъм, за определяне на възраст в групи, базиран на намаляване на размерността, чрез реализирано по метода с ядрото вграждане на възрастовото многообразие в подпространство. Този принос е към Глава 2, точка 2.2.

Приложните приноси са:

1. Програмна реализация и експериментални изследвания на алгоритъм за детектиране на лица, с комбинацията от детектор и класификатор, и последваща регистрация на лица, чрез очният базис. Този принос е към Глави 2,3 и 5, точки 2.2, 3.2, 3.3, 5.1.1, 5.2.1 и 5.2.2.
2. Програмна реализация и експериментални изследвания на алгоритъм за определяне на възраст в групи с Spectral Regression. Постигнат коефициент на разпознаване от 77.67%, използвайки стандартна база данни от лица с анотирани възрасти. Този принос е към Глава 2 и 5, точки 2.2 и 5.1.1.

Алгоритъма за детектиране на лица и алгоритъма за определяне на възраст в групи са докладвани в следната научна публикация:

- [D1] **K. Tonchev**, I. Paliy, O. Boumbarov, S. Sokolov, "Human age-group classification of facial images with subspace projection and support vector machines." *Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems*. Vol. 1. IEEE, 2011. DOI: 10.1109/IDAACS.2011.6072792

3 Разработване на метод и алгоритми за разпознаване на изражения на лица

Човешкото лице е комбинация от физически характеристики и изражения, основаващи се на чувства и емоции. Като такава, то може да бъде използвано за улавяне на вътрешното емоционално състояние на човека. Редовното наблюдение и оценка на човешките изражения може да се използва за определяне на благосъстоянието на човешкия живот. Възможностите за здравни грижи също са огромни, например като виртуален психолог, който помага при диагностицирането на депресия, анализирайки емоциите или детектирането на промени в настроението докато болестта прогресира. В друго поле на приложение, маркетинговият специалист може по-добре да прецени как зрителите отговарят на техните продукти и реклами, като наблюдават израженията им. Учителите също биха могли да преценят дали един урок е интересен на учениците чрез проследяване на вниманието и тяхната реакция изразена чрез лицето. Скорошното развитие на интелигентните автомобили се възползва и от разпознаването на изражението на лицето. Например, интелигентният модул на автомобила може да промени посоката на движение или да го спре, ако възприема, че шофьорът е разстроен, объркан или ядосан. Емоционалното разпознаване може да бъде нова технологична основа за изграждане на системи за подобряване на живота на възрастните хора чрез осигуряване на домашна автоматизация и обща помощ за интелигентни алгоритми. Оценяването на емоционалното състояние на възрастните хора може да бъде важен компонент в тези системи. Честото лошо настроение може да се използва и като сигнал, възрастният да предприеме допълнителни действия за стимулиране на повече дейности, които да подобрят настроението, като например стимулиране на по-активно социално поведение, слушане на музика и т.н. Задачата за разпознаване на изражението на лицето е да имитира човешката визуална система, използвайки камера и компютърни системи или с други думи, за да позволи на машината да има емоционална интелигентност. Това е много предизвикателна задача в областта на компютърното зрение, тъй като изисква не само ефективни техники за анализ на изображения или видео, но също така и подходящи функции, използвани в процеса на машинно обучение.

Разпознаването на изражения е задача от класификацията, при която се цели определянето на принадлежността на изражението на лицето в няколко предварително зададени класа. Класовете изражения обикновено се дефинират в 7 класа: гняв, страх, щастие, отвращение, тъга, изненада, презрение [115]. Тези изражения са идентифицирани като основни изражения на човешкото лице, участващи в социалната комуникация, и поради това се използват в научните изследвания

[116]. Също така, често се използва и неутралното изражение когато се прави сравнителен анализ в разпознаването между отделните изражения и неутралното. В текущата глава са предложени два алгоритъма за разпознаване на изражения на лице на база на изображения на лица. Първият алгоритъм стъпва на метод за избор на признаци, базиран на линейни статистически трансформации, в устойчиво по осветеност log-Gabor признаково пространство. Вторият алгоритъм използва ключови точки от лицето, представляващи геометрията му и прилага обучение на логистична регресия, реализирана по метода с ядрото, но адаптирана за работа с големи обеми от данни.

3.1 Анализ на литературните данни по разпознаване на изражения на лица

С бума на интерактивните системи човек-компютър, се публикуват много научни статии, предлагащи различни подходи за постигане на разпознаване на изражения с различни нива на успеваемост [2]. Сравнително скорошно литературно проучване на алгоритми за разпознаване на изражения е представено в [117]. В проучването са разгледани различните методи за извличане на локални признаци както и геометрични признаци, докладвани в литературата. Един от основните изводи от това проучване е, че извличането на локални признаци, които са инвариантни по отношение на поза и осветление, е от основна важност [118].

В изследването [22], авторите са използвали филтри от типа Gabor и log-Gabor за извличане на признаци и PCA за намаляване на размерността. Класификацията се извършва чрез линеен дискриминантен анализ, използвайки базата данни [119], и постигат точност на разпознаване на израженията от 85%. Предимството на алгоритъма е, че демонстрира ефективността на извличането на признаци с log-Gabor, а като недостатък може да се отчете простотата на класификатора, който се използва.

Авторите в [120] са извършили подробен анализ на алгоритмите за разпознаване на изражения, използвайки Local Binary Patterns (LBP) за извличане на признаци. Тяхното изследване прави сравнителен анализ използвайки различни класификатори, като AdaBoost и SVM. Те отбелязват, че LBP се представя малко по-добре в сравнение с Gabor филтрите, като метод за извличане на признаци, когато анализът се извършва върху изображения с висока разделителна способност. Интересен извод от това проучване е, че когато изображенията са с ниска разделителна способност, филтрите на Gabor дават по-добър резултат. За да анализират ефективността на алгоритмите, те използват базата данни JAFFE [116]. Недостатъкът на това изследване е липсата на сравнение на LBP с други алгоритми за извличане на локални признаци.

В публикацията Ou et al. [121] е предложен алгоритъм, използващ локална информация от изображението на лицето и филтри на Gabor. Те прилагат филтриране на Gabor в околност на 28 ключови точки на лицето, разположени на места като вежди, нос, брадичка и др. За класификация на израженията използват класификатор от типа k-NN, като данните предварително са с намалена размерност чрез прилагане на PCA. Постигнатата степен на разпознаване е 80%, използвайки

базата данни от изражения на лица Cohn-Kanade [122]. Предимство на алгоритъма е, че извлича признаци в околности на ключови точки на лицето като по този начин, извлича информация, която е съществена за задачата. Недостатък на алгоритъма е използването на k-NN за класификация на израженията.

Авторите на [123] предлагат хибридна система, базирана на метода на разреденото представяне, използвайки Gabor уейвлетни филтри за извличане на признаци и SVM класификатор за класификация. Разреденото представяне го постигат чрез апроксимация на тестовите данни с разредена комбинация от речник, построен от обучаващата извадка с Gabor уейвлетите. Техният алгоритъм достига 89% точност, използвайки базата данни с изражения на лица JAFFE [116] и демонстрират ефективността на алгоритъма като го сравняват с други алгоритми, които използват Gabor уейвлети. Основен недостатък е липсата на сравнение на предложеният алгоритъм за разредено представяне с други, които са предложени в литературата.

В работата [124], авторите сравняват степента на разпознаване на израженията на три различни подхода за избор на признаци и обобщаващ подход, базиран на генетичния алгоритъм за оптимизация. Признаците са извлечени с помощта на log-Gabor филтри, а класификацията се извършва с наивен бейсов класификатор. Основен недостатък на алгоритъма е използването на бейсовият класификатор, който в общият случай построява квадратични повърхнини и не може да се справи с по-сложни проблеми за класификация.

Същите автори предлагат и друг подход за разпознаване на изражения, използвайки log-Gabor филтри и избор на признаци в работата [125]. Те адресират високата размерност генерирана от извличането на признаци като използват взаимната информация между съответното изражение и признаковия вектор. Броят на признаците е предварително избран, което е основен недостатък на алгоритъма, защото липсва обективен критерий за избор на броя на признаци. Друг недостатък е използването на наивен бейсов класификатор. Авторите използват Cohn-Kanade и JAFFE базите данни за анализ на ефективността на алгоритъма, но не го сравняват с други алгоритми докладвани в литературата.

Авторите Mehta и Jadhav [126] предлагат алгоритъм за разпознаване на изражението на лицето чрез извличане на признаци с log-Gabor филтър с 5 скали и 8 различни ориентации, последвани от стъпка за намаляване на размерността чрез PCA. Класификацията я извършват използвайки класификатор с евклидово разстояние, което е основен недостатък на алгоритъмът. Тестовите извършват с базата данни JAFFE и постигат средно ниво на разпознаване от 96%. Основен недостатък на този алгоритъм е използването на класификатор с евклидово разстояние, който не взема предвид характеристиките на многообразието на данните.

Повечето бази данни, съдържащи изображения на лица с анотирани изражения са в контролирани условия или такива, които симулират естествени сцени [122]. Анализът на израженията включва много аспекти, които оказват влияние като поза, вариации в осветлението и дори качеството на изображенията, както е описано в проучването [127]. Според това проучване също така, геометрията и/или вида на лицето оказват важна роля при разпознаването на изражения. Адресирането на тази задача за реални изображения, които не са заснети в контролирани условия, изисква използването на големи обеми от данни. Една такава база данни от лица е предложена в [128] и наименувана EmotiW. Използвайки тази база от

данни, в работата [129] е предложен алгоритъм за разпознаване на изражения, базиран на комбинация от конволюционна и рекурентна невронни мрежи. Целта им с първата е да извлекат признаци, които са свързани с промените на лицето в пространството, докато с втората се анализират промените на лицето във времето. Рекурентната мрежа е реализирана с активационни функции Rectified Linear Unit (ReLU), а инициализацията се извършва с специално построена диагонална матрица. Конволюционната мрежа се състои от три слоя, всеки от които използва 8×8 конволюционни маски. Тестовите на алгоритъма с базата данни EmotiW постигат 52.8% точност. Основен недостатък на разработката, е липсата на обосновка на вариацията на резултатите.

Друга работа, използваща базата данни EmotiW е [130]. Авторите ѝ предлагат използването на локалните признаци LBP (виж т. 1.2.1). Признаците след това се трансформират в метрично пространство чрез алгоритъма Multi Dimensional Scaling (MDS) [131], като целта на тази трансформация е да се използва функция за сходство между отделните елементи на данните. След това използват няколко архитектури от конволюционни невронни мрежи и комбинират решението им, с което докладват точност на разпознаване от 54.5%. Недостатък на алгоритъма е липсата на аргументирано използване на комбинацията от различните архитектури на невронните мрежи.

3.1.1 Изводи от литературния анализ по разпознаване на изражения на лица

Изводите от анализа на литературата по определяне на изражения на лица могат да се обобщят до следните:

- Разпознаването на изражения на лица основно се изразява в анализ на геометричните характеристики на лицето, включително и промени, които са на микро ниво.
- Точността на разпознаване на израженията силно зависи от качеството на заснемане на изображението, неговата разделителна способност и потенциални артефакти, възникнали поради голяма компресия. Такъв тип изкривявания влияят върху анализа на геометрията на лицето и най-вече върху микро промените.
- Разпознаването на изражения зависи от вариациите в осветеността, защото по този начин се модулира геометрията на лицето и съответно, е необходимо да се използват признаци които са устойчиви спрямо големи вариации в осветеността.
- За да се адресират всички вариации, касаещи израженията на лицата, е необходимо да се използват големи извадки от данни, съдържащи вариации в осветлението, позата, и качеството, за да се обучат алгоритмите на устойчивост към тези вариации.

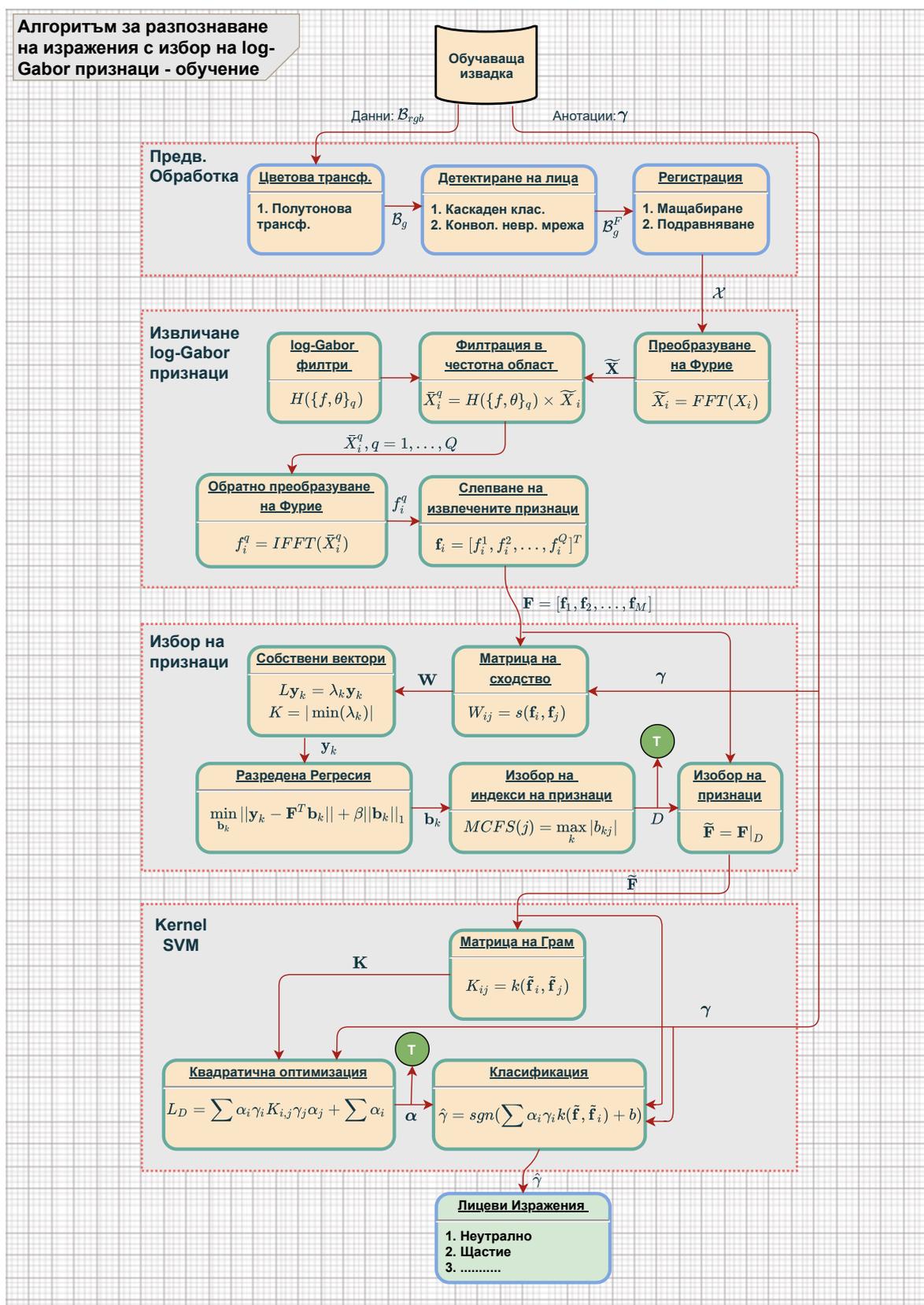
3.2 Разпознаване на изражения на лица чрез избор на признаци в log-Gabor признаково пространство

В тази точка е представен алгоритъм, който класифицира израженията на лице, съдържащо се в изображение на естествена сцена. Алгоритъма се базира на ново-предложена комбинация от алгоритъм за извличане на log-Gabor признаци, и алгоритъм за избор на признаци, който използва анализ на многообразието. Основното предимство на предложеният алгоритъм, спрямо останалите предложени в литературата, е *въвеждането на избор на признаци с което не само се намалява размерността на log-Gabor признаците, но и се запазва геометрията на многообразието на изображенията на лицата така, че да подпомогне и дискриминативността*. Намаляването на размерността също така е важно за класификацията, която е последна стъпка от обработката. В основата си алгоритъма за извличане на признаци се базира на линейните статистически преобразувания, което е свързано с текущия труд.

3.2.1 Блок-схема на потока на информацията и анализ на предложения алгоритъм

Обработката на изображение на естествена сцена, съдържащо лице, минава през няколко етапа и отговаря на общия подход за разпознаване на изображения на обекти представен в т. 1.1.2. Алгоритъмът използва обучаваща извадка за да изчисли коефициентите на операторите, поради което той се реализира в две фази: Обучение и Тест.

На Фиг. 3.1 е представена блок-схема на фазата Обучение на алгоритъма. Първата стъпка е наименувана **Предварителна Обработка**, която основна цел е да подготви изображението за извличането на признаци. Тъй като цвета не е от значение при разпознаването на изражения, а само геометрията, изображението се преобразува в полутоново, за да се намали обема информация необходима за обработка, а и също така, да се подготви за детектора на лица. Детекторът на лица открива лицето в изображение на естествена сцена и предоставя координатите на правоъгълника, който загражда лицето. Използвайки тези координати лицето се изрязва от сцената и се регистрира в обща координатна система. Тази стъпка се нарича *Регистрация* и е цели да се намалят вариациите между позициите и големините на лицата, които са особено важни при холистичен анализ на лица. Обработките, които се прилагат в стъпката **Предварителна Обработка** са важни, за да се намалят неинформативните вариации спрямо задачата, както беше установено в анализа на литературата в предходната точка. Те са идентични с тези, представени в Глава 2, т. 2.2.1. След регистрацията, изображенията на лицата са оразмерени до един и същи размер, и се формират като множество от изображения $\mathcal{X} = \{X_i\}_{i=1}^M$. На това множество съответства вектор γ с дължина M , като всеки елемент от този вектор съдържа анотацията на изражението на лицето, на което съответства индекса на елемента от \mathcal{X} . Така форматирани изображенията се предават на стъпката **Извличане на log-Gabor признаци**.



ФИГУРА 3.1: Блок-схема на алгоритъм за разпознаване на изражения на лица с избор на log-Gabor признаци - фаза Обучение.

В тази стъпка се извличат признаци, използвайки конволюция, но реализирана в честотна област, като изображенията се обработват едно по едно в двумерната област. Първоначално се генерира множество от филтри $H(\{f, \theta\}_q)$, представени чрез двумерни маски, които се прилагат в честотна област. Броят на тези филтри се избира предварително спрямо задачата, която се решава. След като са генерирани филтрите, на всяко едно изображение от \mathcal{X} се прилага двумерното *Преобразуване на Фурие*, при което изображението се трансформира в честотна област където се прилагат маските $\bar{X}_i^q = H(\{f, \theta\}_q) \times \tilde{X}_i$, за да се извърши филтрация. При тази операция от едно изображение се генерират Q на брой матрици, колкото е броят на генерираните филтри, които са със същата размерност, както на изображението. Всяка една от тези матрици се преобразува обратно в областта на изображенията чрез двумерно *Обратното преобразуване на Фурие*. Така на едно изображение X_i се генерират Q на брой изображения f_i^q , всяко едно от които е филтрирано с log-Gabor филтър със съответните параметри. Извлечените признаци се слепват във вектор $\mathbf{f}_i = [f_i^1, f_i^2, \dots, f_i^Q]^T$, а всеки един от тези вектори, съответстващ на изображението X_i , се слепва в матрицата $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_M]$, с което етапа по извличане на log-Gabor признаци е завършен. Мотивацията за използването на този тип признаци е възможността за локализиран анализ, който позволява реализация на различни скали и ориентации. Логаритмичната функция, модулираща честотната характеристика на Gabor уейвлетите, намалява високите честоти и съответно ефекта на шума и осветеността [132]. Изборът на този алгоритъм отговаря и на извода направен в предната точка за избор на признаци, които са устойчиви към вариации в осветеността.

Матрицата с извлечените признаци \mathbf{F} е с размерност $(NQ) \times M$, където Q е броят на приложените филтри. Това означава, че размерността на извлечените признаци е много по-голяма от размерността на изображението, което предполага лоша ефективност заради „проклятието на размерността”. За да се намали размерността се прилага етапа **Избор на признаци** от предложения алгоритъм. Тази стъпка използва концепциите на SR описани в т. 2.2.5, но вместо да извършва линейно преобразуване на признаците, се извършва нелинейно такова. То се състои в намирането на индексите на признаците, които са най-подходящи при вграждането на многообразието им. Първата стъпка е построяването на тегловната матрица \mathbf{W} , която описва сходството между всеки два признакови вектора \mathbf{f}_i и \mathbf{f}_j . Използвайки тази матрица се построява матрицата graph-Laplacian \mathbf{L} и се намират собствените ѝ вектори \mathbf{y}_k , и съответните собствени стойности λ_k . От собствените вектори се избират първите K на брой, които съответстват на първите K собствените стойности с най-малка стойност. По този начин се избират клъстерите, които са най-информативни в пространството в което се вгражда многообразието [114]. Използвайки тези собствени вектори се прилага стъпката *Разредена регресия*, при която се търси разредена линейна комбинация от всички признаци, най-добре описваща вграждането на многообразието. Разредеността се постига чрез L_1 регуляризация на вектора на коефициентите \mathbf{b}_k . След това се преминава към *Избор на индекси на признаци*, при която стъпка се избират $d < N$ на брой индекси D , с което се намалява размерността. Тези индекси се запазват за етапа Тест (обозначено с T заградено в кръг) и се прилагат, за да се изберат d признака $\tilde{\mathbf{F}} = \mathbf{F}|_D$. Мотивацията за използването на този алгоритъм за избор на признаци е следната:

- Алгоритъмът позволява избор на признаци, които запазват дискриминативността на вграденото многообразие на данните.
- Съществуват ефективни алгоритми за оптимизация, с които може точно да се контролира кардиналността на разредения вектор \mathbf{b}_k .
- Изборът на признаци е за предпочитане в случаите, когато извлечените признаци са с по-голяма размерност от тази на изображението, което предполага пре-определеност на базиса на тези признаци.
- В конкретният случай с използваните \log -Gabor признаци, линейна комбинация от тях е нежелателна, защото ще доведе до загуба на локализация и ще увеличи резултатната честотна лента. Поради тази причина изборът на признаци е по-подходящ подход.

Избраните признаци и съответните им анотации се използват за обучението на класификатора SVM, който е описан в Глава 2, т. 2.2.6. Опорните вектори α , изчислени във фазата на обучение, се запазват и използват във фазата Тест.

След като е обучен алгоритъмът се преминава към фазата Тест, при която се използва тестова извадка с известни анотации. Целта на теста е да се определи ефективността на алгоритъмът, след като е бил обучен. Блок-схема на последователността на обработка на информацията във фазата Тест е представена на Фиг. 3.2.

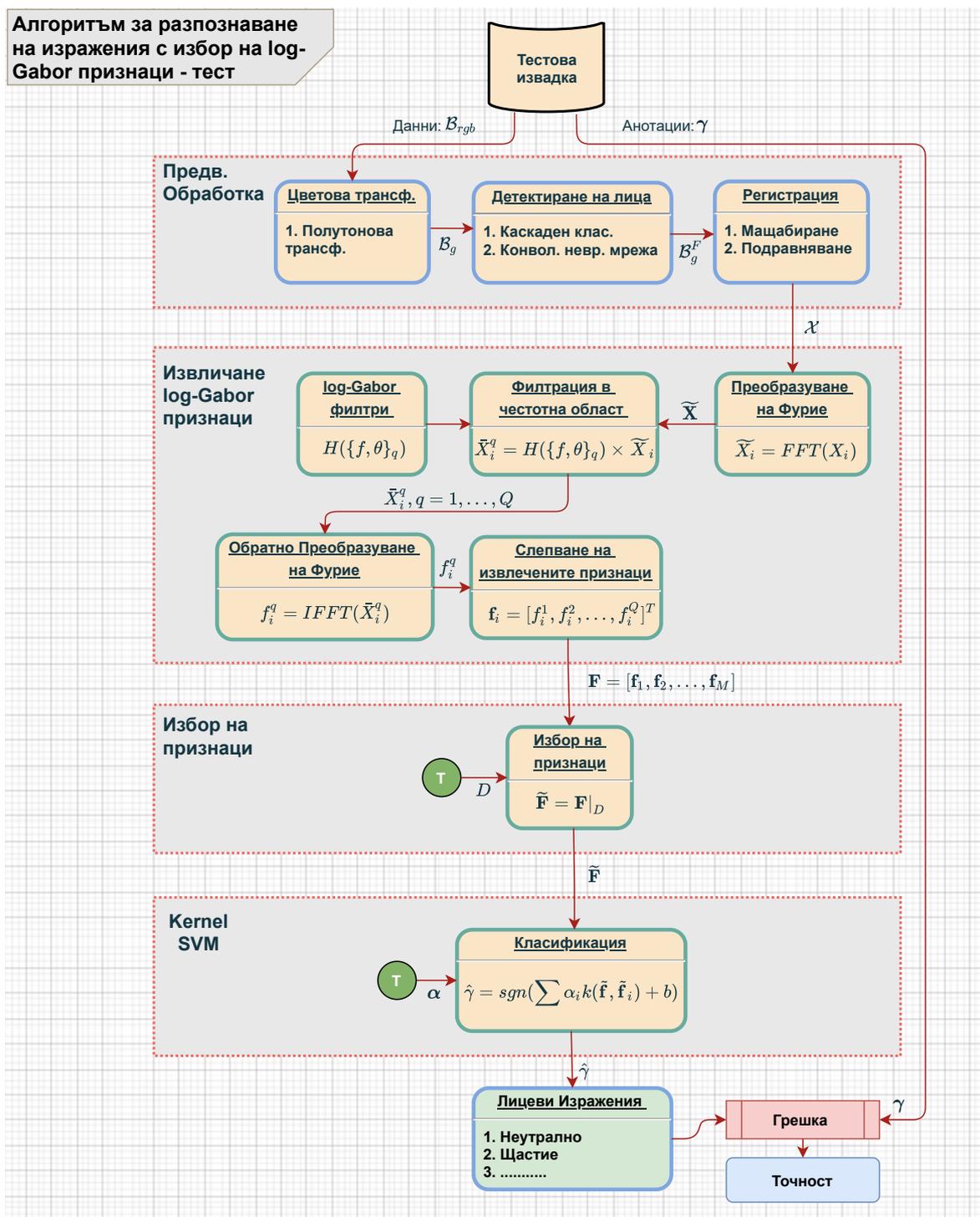
Етапите на **Предварителна Обработка** и **Извличане на \log -Gabor признаци** са идентични както във фазата на обучение. При стъпката **Избор на признаци** се използват индексите на избраните признаци D , изчислени във фазата на обучението, за да се изберат d на брой признака. Избраните признаци се подават към SVM класификатора който използва опорните вектори α от фазата Обучение за да определи типа изражение $\hat{\gamma}$. Ако определеното изражение съвпада с анотацията, то грешката се счита за 1, ако не съвпада, грешката се счита за 0 (блок *Грешка*). Акумулирането на всички грешки и разделянето им на броя на изображенията в тестовата извадка задава точността на алгоритъмът, което е и неговата ефективност (блок *Точност*). Понякога, ефективността се задава в проценти като се умножи по 100.

Определеният тип изражение се сравнява с известната анотацията γ за да се определи точността на алгоритъмът.

3.2.2 Извличане на признаци с \log -Gabor филтри

Израженията на лицето предизвикват специфични промени във формата му, включително поява на бръчки и видима промяна на геометрията на различни части, като например уста и вежди. За да бъдат напълно детектирани тези промени е необходимо да се направи анализ на различни скали, включващ и локална информация.

Често срещан избор за получаване локализирана честотна информация на визуалните модели е да се приложат филтри (уейвлети) на Gabor [133]. Те, обаче, имат два основни недостатъка: максималната широчина на честотната лента е ограничена до приблизително една октава и те не са оптимални в случая, когато се търси широк спектър с максимална пространствена локализация. Известна



ФИГУРА 3.2: Блок-схема на алгоритъм за разпознаване на изразения на лица с избор на log-Gabor признаци - фаза Тест.



ФИГУРА 3.3: log-Gabor филтър в честотна област.

алтернатива на филтъра на Gabor, която адресира тези недостатъци, е наименуванa log-Gabor и е предложена от Field [134]. Тази алтернатива има две важни характеристики. Първата е, че по дефиниция log-Gabor филтрите нямат постоянна съставка. За разлика от тях, Gabor филтрите не могат да бъдат конструирани с произволно широка честотна лента, като същевременно се поддържа сравнително малка постоянна съставка. Втората е, че предавателната функция на log-Gabor филтрите имат добра чувствителност при високите честоти. Математическите означения и последователността на стъпките описани в тази точка отговарят на тези на Фиг. 3.1.

В честотната област log-Gabor филтрите се представят от следната предавателна функция $H(f, \theta)$:

$$H(f, \theta) = H_f \times H_\theta, \quad (3.1)$$

където H_f е радиалната компонента с която се контролира честотната лента на филтъра, а H_θ е ъгловата компонента, която контролира пространствената ориентация на филтъра. Използвайки полярните координати, предавателната функция $H(f, \theta)$ може да бъде изразена и като:

$$H(f, \theta) = \left\{ \frac{-[\ln(\frac{f}{f_0})]^2}{2[\ln(\frac{\sigma_f}{f_0})]^2} \right\} \exp \left\{ \frac{-(\theta - \theta_0)}{2\sigma_\theta^2} \right\}, \quad (3.2)$$

където f_0 е централната честота на филтъра, а θ_0 е ориентацията на филтъра. Параметърa σ_f дефинира радиалната широчина на честотната лента B в октави:

$$B = 2\sqrt{2/\ln 2} \times \left\| \ln\left(\frac{\sigma_f}{f_0}\right) \right\|, \quad (3.3)$$

а параметърът σ_θ дефинира ъгловата лента $\Delta\Omega$ в радиани:

$$\Delta\Omega = 2\sigma_\theta\sqrt{2/\ln 2}. \quad (3.4)$$

На Фиг. 3.3 е изобразено визуално представяне на примерни log-Gabor филтри в честотната област. На фигурите 3.3а и 3.3б са показани двата компонента на филтъра, радиалната и ъгловата, съответно, за произволен мащаб и ориентация. На Фиг. 3.3в е представен резултантният филтър, съчетаващ двата компонента, което се формира като поелементно умножение на отделните филтри.

За извличането на признаци в предложения алгоритъм се използва набор от

log-Gabor филтри в няколко скали и ориентации, с цел да се извлекат черти на лицето, подходящи за разпознаване на израженията. Конволюцията се извършва в честотната област, както беше упоменато, което е значително по-бързо в сравнение с конволюцията в пространствената област.

3.2.3 Избор на признаци чрез разредена регресия

Извличането на признаци цели те да бъдат подходящи за задачата, което може значително да увеличи размерите на данните, както е в случаят с log-Gabor признаците в предложения алгоритъм. Това неизбежно ще доведе до подчертаване на нежелания ефект „проклятието на размерността”, както и ще увеличи изчисленията, които класификаторът трябва да направи. Например, ако входното лице е с размер 100×100 пиксела и избраните скали на log-Gabor филтри са 5, а избраните ориентации 4, тогава получената размерност на признаците е $5 \times 4 \times 100 \times 100 = 200000$, което е значително увеличение на размерността. В този случай, дори и при малки извадки (например 1000 изображения), „проклятието на размерността” е особено изразено и ефективността на класификатора се очаква да падне значително. За да се избегне това нарастване на размерността, обикновено се прилагат два подхода. Първият е чрез намаляване на размерността, при който се търси линейна или нелинейна комбинация от признаци (виж Глава 1). Вторият е избор на признаци, при който се избира малка част от най-представителните признаци, спрямо задачата, която се решава. Това може да се тълкува и като линейно намаляване на размерността чрез разредени линейни комбинации, тоест линейна комбинация, при която голяма част от елементите са нули останалите са единици. Тъй като log-Gabor признаците са реализирани в няколко скали, линейна комбинация от тях ще доведе до увеличаване на резултантната честотна лента, което е в противоречие с реализираните скали. Поради това, изборът на скали и ориентации, които са най-информативни, е по-правилен в сравнение с тяхната линейна комбинация. Математическите означения в тази точка са идентични с тези на Фиг. 3.1 и фигурата може да бъде използвана за референция.

Избраният метод за избор на признаци е предложен в [135], и се базира на вграждането на многообразия или по-точно, вграждането на техните представяния чрез графи. Този подход моделира зависимостта на данните под формата на граф, намиращ се в пространството на данните. Целта е вграждането на този граф в пространство с по-ниска размерност, като същевременно се запазват зависимостите на данните. С други думи, целта е да се избере пространство с минимални размери, което запазва зависимостите на данните. Този подход следва общата рамка за моделиране на линейните статистически преобразувания, описани в т. 1.2.6. При този подход, зависимостите между данните се моделират чрез граф, при който всеки ръб на графа има асоциирано тегло, което в общият случай представлява някакво сходство между два елемента от данните. Тези зависимости се описват чрез тегловна матрица \mathbf{W} , където всеки елемент W_{ij} кодира връзката между данните i и j . Построяването на тази матрица, използвана в алгоритъма за избор на признаци е идентично с това, описано в т. 2.2.5, но за пълнота ще бъде повторено тук. Всеки един от елементите на \mathbf{W} приема една от следните стойности:

- $W_{i,j} = 1/l_k$ ако $\tilde{\mathbf{x}}_i$ и $\tilde{\mathbf{x}}_j$ заедно принадлежат на k -тия клас, където l_k е броят на елементите в него;
- $W_{i,j} = \delta s(i, j) = \delta \exp\left(-\frac{\|\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j\|}{\sigma}\right)$, ако $\tilde{\mathbf{x}}_i$ е сред трите най-близки съседи на $\tilde{\mathbf{x}}_j$ или обратно. Функцията $s(i, j)$ изчислява сходството между $\tilde{\mathbf{x}}_i$ и $\tilde{\mathbf{x}}_j$, за $\sigma \in \mathbb{R}$. Параметърът δ се използва за настройка на теглото между елементите които имат етикети и тези които имат голямо сходство изчислено чрез $s(i, j)$. Обикновено $\delta = 0.1$, а σ се изчислява експериментално;
- $W_{i,j} = 0$ във всички останали случаи.

По този начин се задава непълно свързан граф, на който всеки от ръбовете е асоцииран с тегло. Теглата са избрани така, че да се запази дискриминативността, но и също така да се запази геометричната структура на данните при вграждането на графа в пространството с по-ниска размерност.

С тегловната матрица се изчислява graph-Laplacian матрицата по следният начин: $\mathbf{L} = \mathbf{D} - \mathbf{W}$, където $D_{jj} = \sum_i W_{ij}$. Диагонализацията на матрицата чрез задачата за собствените стойности и вектори предоставя вграждането в пространството с по-ниска размерност. Тоест, собствените вектори на този оператор осигуряват оптимално вграждане на графа, запазвайки съседството на точките данни в пространството на вграждане. Както е показано на Фиг. 3.1, собствените стойности и вектори на L се изчисляват с:

$$L\mathbf{y}_k = \lambda_k \mathbf{y}_k, \quad (3.5)$$

където \mathbf{y}_k са собствените вектори, които осигуряват вграждането на графа. От всички собствени вектори се избират K на брой, където $K < M$, които запазват клъстерната структура на многообразието. На Фиг. 3.1 този избор е изразен чрез $K = |\min(\lambda_k)|$.

Избраните собствени вектори се използват за линеаризация на функцията на вграждане, но чрез разреждени решения използвайки L_1 нормата. Този подход е валиден и е идентичен с алгоритъма SR, описан в т. 2.2.5. Тези решения се намират чрез решаване на задачата за регресия:

$$\min_{\mathbf{b}_k} \|\mathbf{y}_k - \mathbf{F}^T \mathbf{b}_k\| + \beta \|\mathbf{b}_k\|_1 \quad (3.6)$$

където \mathbf{F} са извлечените log-Gabor признаци. Резултатът при решаването на тази задача е набор от K разреждени вектори \mathbf{b}_k , всеки от които представлява разредената комбинация от признакови вектори, осигуряващи най-доброто вграждане на графа, представено от собствените вектори. Решаването на тази задача включва избора на параметъра β , с който се контролира разредеността на решението \mathbf{b}_k , тоест броя на ненулевите елементи. Изборът му по принцип се базира на експерименти, но в текущият случай, формулировката на оптимизационната задача позволява да се използва алгоритъма Least Angle Regression (LAR) [136]. Този алгоритъм допуска кардиналността на решенията да бъде зададена в явен вид, тоест борят на ненулевите елементи, и поради това му предимство, той се използва в тази работа за решаване на (3.6). При използването на алгоритъма, кардиналността

на решенията се избира да бъде $d = |\gamma| + 1$, тоест броят на класовете $+1$, за да се запази дискриминативността.

Така намерените разредени решения $\{\mathbf{b}_k\}_{k=1}^K$ сами по себе си не могат да бъдат директно използвани за избора на признаци. Въпреки, че те са с кардиналност d , те са K на брой и е необходимо допълнително правило за избор на само d индекса на признаците. Интуитивно решение, предложено от [135] и наименувано Multi-Cluster Feature Selection (MCFS) е такова, че за всеки индекс на признаците да се избере максималната ненулева стойност от разредените решения:

$$MCFS(j) = \max_k |b_{kj}|, \quad (3.7)$$

където b_{kj} е j -тият елемент на k -тият вектор. MCFS избира най-високото тегло на векторния елемент от всички вектори \mathbf{b}_k . С други думи, MCFS избира най-представителното тегло за дадена функция. След този процес на селекция следващата стъпка е да се подредят векторни компоненти и съответните им индекси във възходящ ред. Накрая функциите се избират чрез избиране на индексите на първите d водещи компоненти на подредения вектор, а индексите на тези вектори се запазват в множеството D , за да бъдат използвани във фазата Тест. Този процес на избор на функции се извършва по време на обучението и води до избор от индекси на най-представителните log-Gabor признаци.

3.3 Разпознаване на изражения на лица за големи обеми от данни реализиран по метода с ядрото

На база на литературният анализ по разпознаване на изражения на лица, един от важните изводи направени в т. 3.1.1 е, че точността на разпознаването зависи от качеството на заснетите изображения, позата и осветеността. Много често в практиката, обаче, изображенията са с ниско качество или голяма компресия, например в Интернет пространството, а позата е случайна и често в анфас. Това значително усложнява анализа и налага използването на по-ефективни алгоритми, които са с по-голям капацитет на обучение и съответно по-голям обем на обучаващата извадка. Целта на голямата извадка е да обхваща всички вариации на позата на лицето, вариации на осветеността и качество на изображението. Това, обаче, води до нуждата от повече изчислителна мощ, както и по-ефикасни алгоритми, които да могат да работят с такъв обем от данни. Също така, налага използването на алгоритми, които да извличат признаци, свързани с геометрията на лицето, които алгоритми да работят с много висока точност.

В тази точка е описан ново-предложен алгоритъм, който разпознава изражения на лица, използвайки голяма обучаваща извадка. Извличането на признаци на предложения алгоритъм е реализирано с оператор, който детектира ключови точки на лицето с голяма точност, защото се обучава с голяма извадка от данни. Класификацията е реализирана с класификатор използващ метода с ядрото. Както беше установено от изводите в Глава 1, методите с ядро изискват квадратично повече изчисления, зависещи от обема на данните. Това означава, че директната имплементация е невъзможна и е нужен друг подход. Използваният подход е базиран на случайни признаци, с които се апроксимира матрицата на Грам, и което

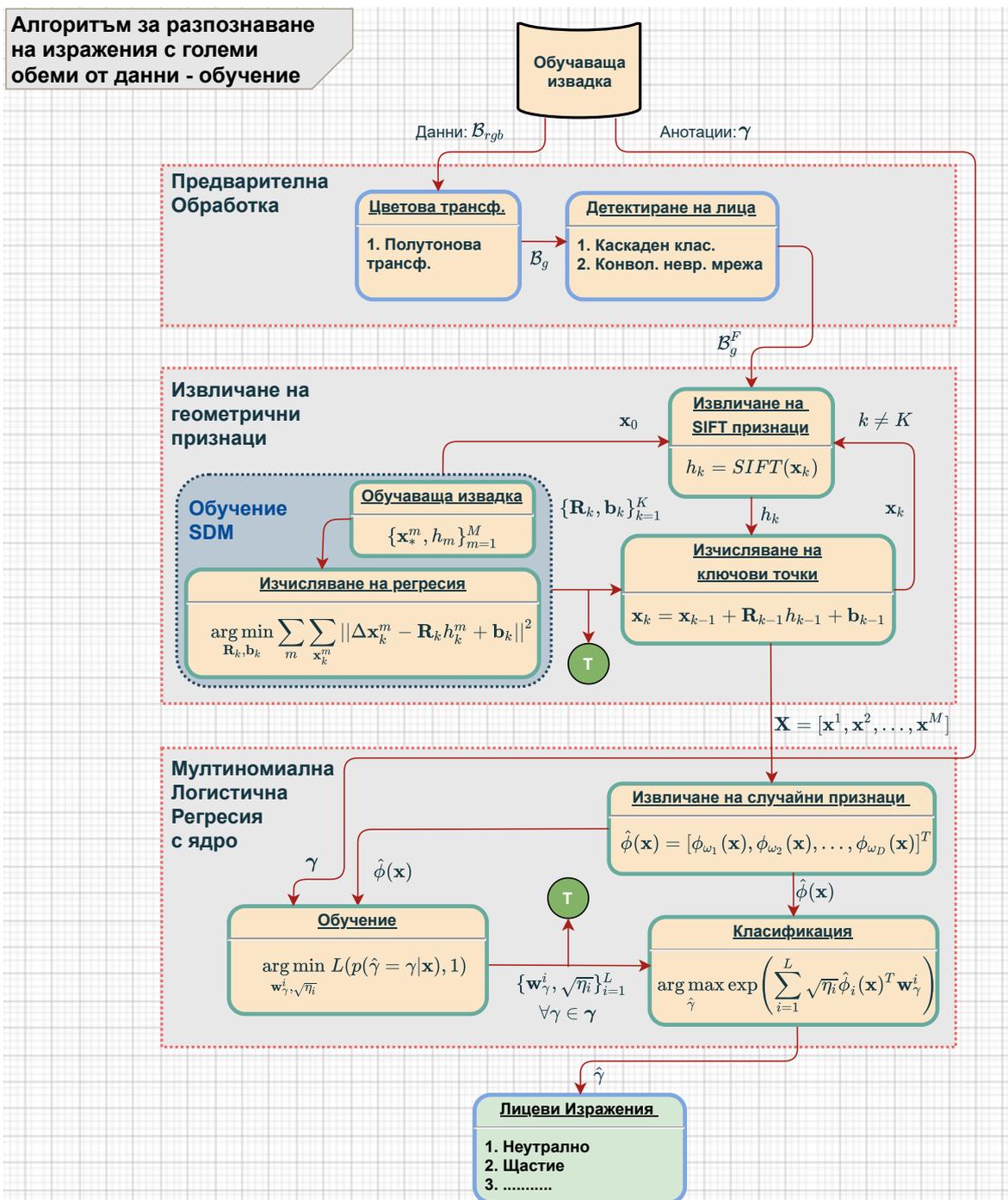
позволява работата с големи обеми от данни. Предимството на предложеният алгоритъм, спрямо другите предложени в литературата, е *използването на точен метод за извличане на геометрични характеристики и използване на метода с ядрото, адаптиран за работа с големи обеми от данни.*

3.3.1 Блок-схема на потока на информацията и анализ на предложения алгоритъм

Предложеният алгоритъм се реализира в две фази, Обучение и Тест, които отговарят на общият подход за разпознаване на изображения на обекти представен в т. 1.1.2. Във фазата Обучение се изчисляват коефициентите на съответните оператори с обучаваща извадка, а във фазата Тест, те се използват за да се определи ефективността на алгоритъма.

На Фиг. 3.4 е представена блок-схема на фазата Обучение на алгоритъма. Първият етап е наименуван **Предварителна Обработка**. В този етап се подготвя изображението за извличането на геометрични признаци, като първоначално, изображението се преобразува в полутоново за да се намали обема на информация необходима за обработка, тъй като цвета не е от значение при разпознаването на изразения. В следващата стъпка детектора на лица открива лицето в изображението и предоставя координатите на правоъгълника, който загражда лицето. В тази стъпка също така се изрязва лицето, спрямо правоъгълника, и се предоставя изображение на изрязаното лице. Обработките които се прилагат в стъпката **Предварителна Обработка** са важни, за да се отстрани фона и да се извърши предварителна регистрация на лицето чрез детектора на лице. Те са идентични с тези, представени в Глава 2, т. 2.2.1.

След регистрацията, изображенията на лицата са оразмерени до един и същи размер и се формират, като множество от изображения $\mathcal{B}_g^F = \{B_i\}_{i=1}^M$. На това множество съответства вектор γ с дължина M , като всеки елемент от този вектор съдържа анотацията на изразението на лицето на което съответства индекса на елемента от \mathcal{B} . Така форматирани, изображенията се предават на стъпката за **Извличане на геометрични признаци**. В тази стъпка се извличат признаци представляващи геометрията на лицето, която е от основно значение за разпознаването на изразения. Това описание се представя под формата на ключови точки разположени на части от лицето, като очи, уста, нос и т.н. При промяна на изразението на лицето, респективно геометрията му, тези точки променят координатите си. Алгоритъмът който се използва за детектиране на ключовите точки е наименуван Supervised Descent Method (SDM) и е предложен в работата [137]. Той детектира общо 49 ключови точки, представени чрез координатите на изображението, използвайки итеративна процедура. Тя се прилага върху извлечени признаци от тип Scale Invariant Features Transform (SIFT) (виж т. 1.2.1) от области на изображението с размери 32×32 пиксела, всяка една от които центрирана върху ключова точка. Итеративната процедура обикновено се прилага в не повече от 5 стъпки, като всяка стъпка се състои в афинна трансформация, което говори за голямо бързодействие на алгоритъма. При инициализацията на алгоритъма се използват ключови точки \mathbf{x}_0 , които са средни за обучаващата извадка, както и съответните на тях SIFT признаци h_k . Обучението на алгоритъма за детектиране на ключовите точки се извършва като се решава квадратична регресия без



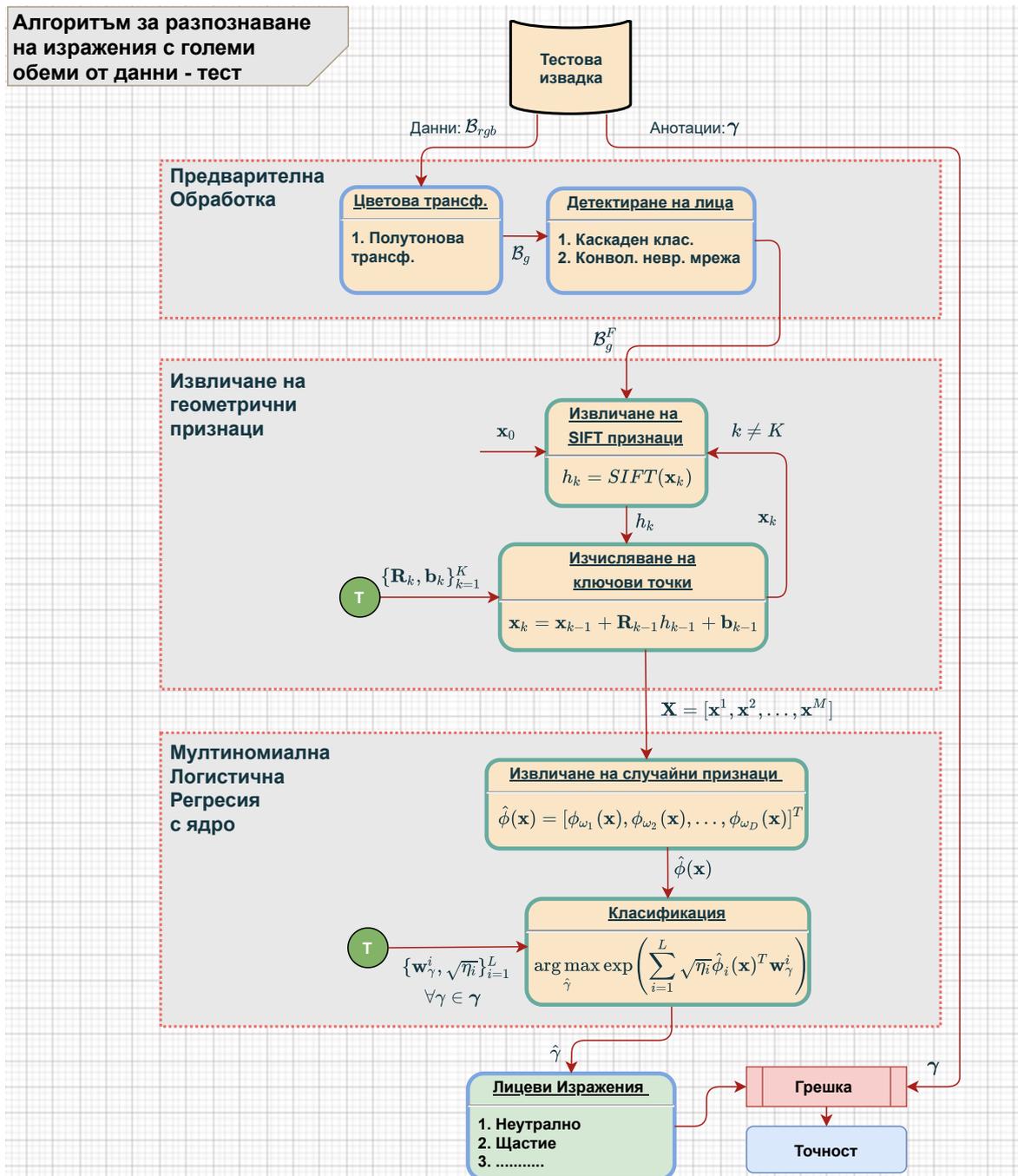
ФИГУРА 3.4: Блок-схема на алгоритъм за разпознаване на изражения на лица с геометрични признаци и метода с ядрото, реализиран за големи обеми от данни - фаза Обучение.

ограничения [137]. Тази процедура обаче се извършва преди самото обучение на предложеният алгоритъм, използвайки друга обучаваща извадка предназначена за детектиране на ключови точки. Коефициентите на итеративната процедурата $\{\mathbf{R}_k, \mathbf{b}_k\}_{k=1}^K$ се запазват и се използват при етапа Тест (обозначено с Т на фигурата). Мотивацията за избора на този алгоритъм за извличане на геометрични признаци се състои в неговата ефективност, устойчивост на вариация в позата и осветлението. С този алгоритъм се удовлетворява изводът направен от литературният анализ, а именно, че геометрията на лицето е основна характеристика при разпознаването на изражения.

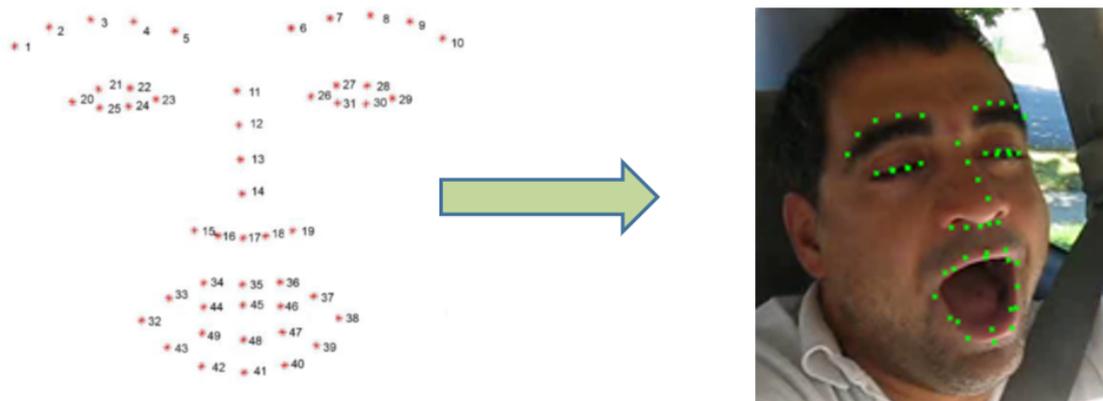
Детектираните ключови точки се подреждат в матрица и се подават към следващата стъпка **Мултиномиална Логистична Регресия с Ядро**, която реализира класификация в няколко класа лицеви изражения. Както самото име подсказва, в тази стъпка класификаторът е реализиран с Мултиномиална Логистична Регресия в която се използват функции, реализирани по метода с ядрото. Характерна особеност на този класификатор е, че той е реализиран за работа с големи обеми от данни. Това е постигнато, използвайки стохастична апроксимация на ядрата, което позволява да се работи с по-малък брой изчисления. Този тип апроксимация също така позволява използването на множество ядра за постигане на по-голям капацитет на обучение. Поради тази причина, подходът с множество ядра е приет в предложеният алгоритъм, използвайки гаусови и лапласови ядра. Първата стъпка от този алгоритъм е *Извличане на случайни признаци* които са $\hat{\phi}(\mathbf{x})$ от всеки един вектор на геометричните признаци \mathbf{X} . Използвайки етикетите и случайните признаци се извършва *Обучение* което се състои в изпъкнала оптимизация, при която се търсят теглата на мултиномиалната регресия $\{\mathbf{w}_\gamma^i\}_{i=1}^L$ и теглата на множеството ядра $\{\sqrt{\eta_i}\}_{i=1}^L$. Обучението се извършва като първо, се изчисляват теглата на регресията, след което се фиксират и се изчисляват теглата на ядрата. Използвайки теглата от обучението се извършва *Класификация*, като се прилага логистичната регресия за всеки един от класовете γ и се избира класовата принадлежност за който стойността е най-голяма. Теглата от обучението се запазват за етапа тест, както е обозначено на фигурата с Т. Мотивацията за избора на този алгоритъм е следната:

- Алгоритъмът позволява използването на големи обеми от данни при реализация на методи, които използват ядро.
- Използването на множество ядра лесно се интегрира в алгоритъма, което позволява използването им и съответно с това се увеличава капацитета на обучение.
- Удовлетворява се изискването за работа с големи обеми от данни, когато е необходимо да се адресират различни вариации (осветеност, поза и т.н.), както беше установено в литературния анализ на разпознаването на изражения.

След като е обучен алгоритъма за разпознаване на изражения се преминава към фазата Тест, при която се използва тестова извадка с известни анотациите. Целта на теста е да се определи точността на разпознаване на израженията от алгоритъма, след като е бил обучен. Блок-схема на последователността на обработка на информацията във фазата Тест е представена на Фиг. 3.5. Етапите **Предварителна Обработка** и **Извличане на геометрични признаци** са идентични



Фигура 3.5: Блок-схема на алгоритъм за разпознаване на изражения на лица с геометрични признаци и метода с ядрото, реализиран за големи обеми от данни - фаза Тест.



ФИГУРА 3.6: Илюстрация, поясняваща целта на метод за детектиране на ключови точки по лицеви признаци.

както във фазата Обучение, като за детектирането на ключови точки се използват същите коефициенти използвани във фазата Обучение. При стъпката **Мултиномиална Логистична Регресия с Ядро** се използват коефициентите на множеството ядра и мултиномиалната регресия, изчислени във фазата Обучение, за да се определи класа на изражението. Ако определеното изражение съвпада с анотацията, то грешката се счита за 1, ако не съвпада, грешката се счита за 0 (блок *Грешка*). Акумулирането на всички грешки и разделянето им на броя на изображенията в тестовата извадка задава точността на алгоритъма, което е и неговата ефективност (блок *Точност*). Понякога, ефективността се задава в проценти, като се умножи по 100.

3.3.2 Извличане на геометрични признаци с Supervised Descent Method

Извличането на геометрични признаци в предложеният алгоритъм се базира на детектирането на ключови точки на лицето, които се смятат за представителни и също така, достатъчно точно описващи геометрията на лицето при промяна на изражението. Изискванията които се поставят по отношение на избора такъв тип метод са свързани с преимуществото му над останалите, по отношение на точността на локализиране на ключовите точки при различна ориентация на главата, неравномерна осветеност, вариация в израженията на човека, възможност за работа при частично закриване на лицето (напр. с ръка или др.) и не на последно място - работа в реално време. На Фиг. 3.6 е демонстрирано нагледно каква е целта на използването на такъв метод, а именно - трансформиране на предварително дефиниран шаблон от точки според изражението на лицето, където в ляво е представен шаблона състоящ се от 49 точки, а в дясно на фигурата, е трансформираната му версия спрямо изображението на лицето, което се анализира. Математическите означения в тази точка отговарят на Фиг. 3.4 и фигурата може да се използва за референция.

В литературата са докладвани различни методи за детектиране на ключови точки на лицето по описания начин. Голяма част от тях са базирани на добре известните: Модели с Активен Шаблон (*ASM*) [138], Активни Модели на Вида

(AAM) [139] и Локални Модели с Ограничения, (Constrained Local Models (CLM)) [140]. По-нови разработки, известни в литературата, са такива базирани на методи с примерни данни [141] и регресионни методи с гласуване [142]. Направеният сравнителен анализ от авторите на алгоритъма наименуван SDM [137], с други алтернативни методи, показва превъзходството му спрямо останалите, както и това, че този метод удовлетворява изискванията за надеждност при големи вариации на поза, осветеност и качество на изображението. Освен това е демонстрирано, че предложеният алгоритъм може да работи в реално време. Тези предпоставки представляват основен мотив при избора на SDM за извличане на геометрични признаци, базирани на ключови точки на лицето.

Алгоритъмът използва обучаваща извадка от база данни на лица с ръчно аномтирани ключови точки. За да се постигне инвариантност по отношение на позата на главата, изражението и вариации в осветеността, е необходимо лицето на даден индивид да бъде заснето от различен ъгъл при различни нива на осветеност, както и при различни изражения. Често това се постига чрез заснемането на субекта с няколко камери, разположени на различни ъгли спрямо субекта. В основата на алгоритъма стои итеративно прилагане на афинна трансформация върху предварително извлечени признаци, които съдържат локална информация. Тези признаци може да са произволни, но авторите на SDM предлагат използването на оператора SIFT (виж т. 1.2.1), който извлича признаци, инвариантни по отношение на мащаба и ориентацията. Като резултат от прилагането на SIFT, извлечените признаци представляват 49 хистограми, съответстващи на всяка една от ключовите точки. Пресмятането на всяка една хистограма се извършва върху област от изображението с размери 32×32 пиксела, центрирана във всяка една от ключовите точки, при което се генерират хистограми с размерност 128 елемента. Използвайки тези признаци, SDM изчислява трансляционни вектори за всяка една от ключовите точки, които описват деформирането на шаблона за подравняване върху лицето което се анализира.

Нека с \mathbf{x}_0 са обозначени точките на шаблона, а с h_0 е отбелязан векторът, който съдържа всички хистограми на SIFT признаците подредени така, че да съответстват на всяка една ключова точка от \mathbf{x}_0 . Размерността на този вектор следователно е $49 \times 128 = 6272$ елемента. Една итерация от алгоритъма SDM се задава с:

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \mathbf{R}_{k-1}h_{k-1} + \mathbf{b}_{k-1}, \quad (3.8)$$

където \mathbf{b}_k е вектор с размерност 98 елемента, а \mathbf{R}_k е матрица с размерност 98×6272 и те се изчисляват през фазата на обучение. Както се вижда от уравнението, това е афинна трансформация която цели да преобразува признаците SIFT в пространството на ключовите точки. Итерационната процедура използва за начало \mathbf{x}_0 и съответните им h_0 . Признаковият вектор h_{k-1} се изчислява като се използват точките, изчислени при съответната итерация \mathbf{x}_{k-1} .

Целта на обучението на алгоритъма е да бъдат изчислени коефициентите на афинната трансформация $\{\mathbf{R}_k, \mathbf{b}_k\}_{k=1}^K$, за съответният брой итерации K , който брой също така се определя по време на обучението. Нека е дадена обучаваща извадка от M на брой данни $\{\mathbf{x}_*^m, h_m\}_{m=1}^M$, където \mathbf{x}_*^m са аномтираните ключови точки за различни изражения на лицето, а h_m са съответните им SIFT признаци. Изчисляването на \mathbf{R}_k и \mathbf{b}_k за итерация k се извършва чрез решаване на следната

задача за регресия:

$$\arg \min_{\mathbf{R}_k, \mathbf{b}_k} \sum_m \sum_{\mathbf{x}_k^m} \|\Delta \mathbf{x}_k^m - \mathbf{R}_k h_k^m + \mathbf{b}_k\|^2, \quad (3.9)$$

където $\Delta \mathbf{x}_k^m = \mathbf{x}_*^m - \mathbf{x}_k^m$, тоест това е отместването изчислените точки спрямо истинските, като изчислените се пресмятат използвайки коефициентите на афинната трансформация от предходната итерация (3.8), и това се извършва за всеки елемент m от обучаващата извадка. Признаците SIFT се преизчисляват за всяка итерация k и изображение m , като около точките \mathbf{x}_k^m се центрира прозорец с размер 32×32 пиксела на който се изчисляват SIFT признаците. Усредняването се извършва както по отделните изображения от обучаващата извадка m , така и по отделните точки от всяка итерация \mathbf{x}_k^m , с което се цели апроксимация на разпределението на всички начални и междинни позиции на ключови точки \mathbf{x} . Целевата функция на (3.9) е изпъкнала от което следва, че съществува глобален минимум за всяка итерация. Това, че е изпъкнала, позволява използването на стандартни алгоритми за изпъкнала оптимизация, като в случая с голям брой данни се използва Stochastic Gradient Descent (SGD) [143]. Итерациите се прекратяват, когато разликата на грешките между две последователни итерации стане по-малка от предварително зададен праг. Според проведените експерименти в този труд, а и експериментите от авторите на SDM, този брой обикновено е не повече от 5, което само по себе демонстрира ефикасността и ефективността на алгоритъма.

След като бъдат изчислени коефициентите \mathbf{R}_k и \mathbf{b}_k на афинната трансформация, за съответният брой итерации, алгоритъмът SDM се прилага итеративно, използвайки (3.8).

3.3.3 Разпознаване на изражения с метода с ядрото реализиран за големи обеми от данни

Както бе разгледано в Глава 1, методът с ядрото позволява постигане на голям капацитет на обучение чрез имплицитно вграждане в пространство с по-висока или потенциално безкрайна размерност (на теория). Основният недостатък на този метод се явява необходимостта от изчислителни ресурси, която нараства квадратично като функция на размера на обучаващата извадка. Запазването на матрицата на Грам изисква $O(M^2)$ елемента памет и $O(M^2d)$ изчислителни операции (M е размера на обучаваща извадка; d е размерността на данните). Поради тази причина, в задачите с големи обеми от данни, метода с ядрото не се използва в директната му форма. Вместо това се използват апроксимации на метода с ядрото и това е активна тема в научните изследвания. Математическите означения в тази точка отговарят на Фиг. 3.4 и фигурата може да се използва за референция.

Апроксимацията на ядрата със случайни признаци е един от обещаващите подходи за разрешаване на проблема със скалируемостта на методите с ядрото. Поради това в предложеният алгоритъм се използва методът описан в [144]. Основното предимство на този подход се състои в намаляване на времето за обучение и тест на алгоритмите, както и намаляване на необходимата памет. Постигнатата апроксимация е за ядра които са инвариантни по отношение на трансляция, а именно

гаусовото и лапласовото ядро:

$$\begin{aligned} k_G(\mathbf{x}, \mathbf{z}) &= e^{-\|\mathbf{x}-\mathbf{z}\|^2/2\sigma^2}, \\ k_L(\mathbf{x}, \mathbf{z}) &= e^{-|\mathbf{x}-\mathbf{z}|_1/\sigma}. \end{aligned} \quad (3.10)$$

Основният резултат на Rachimi и Recht [145] доказва, че ядрата които са инвариантни по отношение на трансляция, могат да бъдат изразени чрез хармонични функции:

$$k(\mathbf{x} - \mathbf{z}) = \int_{\mathbb{R}^d} p(\boldsymbol{\omega}) e^{j\boldsymbol{\omega}^T(\mathbf{x}-\mathbf{z})} d\boldsymbol{\omega}, \quad (3.11)$$

където d е размерността на данните. При апроксимацията на гаусовите и съответно лапласовите ядра се използват гаусовото и лапласово разпределение:

$$p_G(\boldsymbol{\omega}) = \mathcal{N}(0, \frac{1}{\sigma} \mathbf{I}), \quad p_L(\boldsymbol{\omega}) = \prod_d \frac{1}{\pi(1 + \sigma^2 \omega_d^2)}, \quad (3.12)$$

където \mathbf{I} е идентитета. Използвайки факта, че $p(\boldsymbol{\omega})$ е вероятностна мярка, (3.11) се явява математическото очакване спрямо $\boldsymbol{\omega}$, тоест:

$$k(\mathbf{x} - \mathbf{z}) = \mathbb{E}_{\boldsymbol{\omega}}[e^{j\boldsymbol{\omega}^T \mathbf{x}} e^{-j\boldsymbol{\omega}^T \mathbf{z}}]. \quad (3.13)$$

Това позволява апроксимиране на ядрото като математическото очакване се замени с усредняване на извадка от случайни признаци [144]. По-точно, нека са извлечени D честоти от разпределението $p(\boldsymbol{\omega})$, $\{\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \dots, \boldsymbol{\omega}_D\} \sim p(\boldsymbol{\omega})$, тогава апроксимацията на ядро, инвариантно по отношение на трансляция, може да се изрази като:

$$k(\mathbf{x} - \mathbf{z}) \approx \frac{1}{D} \sum_{i=1}^D \phi_{\boldsymbol{\omega}_i}(\mathbf{x}) \phi_{\boldsymbol{\omega}_i}(\mathbf{z}) = \hat{\boldsymbol{\phi}}(\mathbf{x})^T \hat{\boldsymbol{\phi}}(\mathbf{z}), \quad (3.14)$$

където $\phi_{\boldsymbol{\omega}_i}(\cdot) : d \rightarrow M$ е изображение в хилбертовото пространство с възпроизвеждащо ядро, което е с по-висока размерност ($d \ll M$). Хармоничните функции с които се изразява ядрото в (3.11) са комплексни, но в случая с реални данни, подходящ избор са косинусовите базисни функции:

$$\phi_{\boldsymbol{\omega}_i}(\mathbf{x}) = \sqrt{\frac{2}{D}} \cos(\boldsymbol{\omega}_i^T \mathbf{x} + b_i), \quad (3.15)$$

където b_i е случайна величина, извлечена от равномерно разпределение в интервала $[0, 2\pi]$. Използвайки (3.15), векторното представяне на признаковите вектори $\hat{\boldsymbol{\phi}}(\cdot)$ е както следва:

$$\hat{\boldsymbol{\phi}}(\mathbf{x}) = [\phi_{\boldsymbol{\omega}_1}(\mathbf{x}), \phi_{\boldsymbol{\omega}_2}(\mathbf{x}), \dots, \phi_{\boldsymbol{\omega}_D}(\mathbf{x})]^T. \quad (3.16)$$

Именно това представяне се използва при изчисляването на апроксимацията на ядрото или по-точно, това е апроксимация на признаковите функции които изобразяват от пространството на изображенията (или друго векторно пространство) в хилбертово пространство с възпроизвеждащо ядро. То е особено ефективно, когато размерността D е значително по-малка от броя на данните M .

Това представяне може да се използва във всички методи с ядро, но в тази работа се използва мултиномиална логистична регресия, както е предложено в [144], реализирана по метода предложен в [145]:

$$p(\hat{\gamma} = \gamma | \mathbf{x}) = \frac{e^{\mathbf{w}_\gamma^T \hat{\phi}(\mathbf{x})}}{\sum_c e^{\mathbf{w}_\gamma^T \hat{\phi}(\mathbf{x})}}, \quad (3.17)$$

където $\hat{\gamma}, \gamma \in \gamma$ е класа. Векторът \mathbf{w}_γ се изчислява за съответният клас и представлява линейната комбинация на елементите на признаковия вектор за класа γ . За да използват паралелни изчисления, авторите предлагат да разделят линейната комбинация в (3.17) на няколко блока B . Така, признаковият вектор $\hat{\phi}$ се разделя на няколко отделни под-вектора с размерност B и регресията се изчислява за всеки един от тях паралелно. При това разделяне, мултиномиалната логистична регресия (3.17), реализирани за паралелни изчисления, се изразява с:

$$\begin{aligned} p(\hat{\gamma} = \gamma | \mathbf{x}) &\propto \exp \left(\frac{1}{B} \sum_{i=1}^B \hat{\phi}_b(\mathbf{x})^T \mathbf{w}_\gamma^b \right) \\ &= \sqrt{\prod_b \exp(\hat{\phi}_b(\mathbf{x})^T \mathbf{w}_\gamma^b)}. \end{aligned} \quad (3.18)$$

Намиране на параметрите на логистичната регресия е изпъкнала оптимизационна задача без ограничения, което означава, че е достатъчно да се използва алгоритъма за спускане по градиента. Поради голямата обучаваща извадка обаче, е необходимо да се използват стохастични методи, като авторите на [144] предлагат използването на Stochastic Averaged Gradient (SAG) [146], който е специално разработен за изпъкнали задачи. Този алгоритъм се използва и в текущата работа.

Изборът на ядро е една от важните предварителни стъпки при реализацията на този алгоритъм. Използването на комбинация от много ядра обаче е за предпочитане, защото осигуряват много по-голям капацитет на обучение [147]. Поради това, вместо едно ядро в (3.17) се използва конична комбинация от гаусови и лапласови ядра. Нека са дадени L базови ядра $\{k_i(\cdot, \cdot)\}_{i=1}^L$, тогава коничната им комбинация се изразява чрез:

$$k(\cdot, \cdot) = \sum_{i=1}^L \eta_i k_i(\cdot, \cdot), \quad \eta_i \geq 0 \quad (3.19)$$

където $k(\cdot, \cdot)$ е резултантното ядро. Тогава във фазата Обучение, решаването на логистичната регресия се свежда и до намиране на коефициентите на комбинацията от ядра η_i . Използвайки представянето чрез случайни признаци (3.16), коничната комбинация от ядра се представя по следният начин:

$$k(\cdot, \cdot) \approx \sum_{i=1}^L \eta_i \hat{\phi}_i(\cdot)^T \hat{\phi}_i(\cdot) = \hat{\phi}(\cdot)^T \hat{\phi}(\cdot), \quad (3.20)$$

където $\hat{\phi}(\cdot)$ са скалирани с $\sqrt{\eta_i}$ и са матрици с размер $L \times D$. Използването на тази апроксимация в мултиномиалната регресия става като първо, за всеки $\{\hat{\phi}_i\}_{i=1}^L$ и

всеки клас γ , се намират коефициентите на мултиномиалната регресия $\{\mathbf{w}_\gamma^i\}_{i=1}^L$, след което те се фиксират и се изчисляват коефициентите на коничната комбинация:

$$p(\hat{\gamma} = \gamma | \mathbf{x}) \propto \exp \left(\sum_{i=1}^L \sqrt{\eta_i} \hat{\phi}_i(\mathbf{x})^T \mathbf{w}_\gamma^i \right). \quad (3.21)$$

Тази задача е също изпъкнала оптимизационна задача което позволява решаването и с SAG.

С изчисляването на теглата $\{\mathbf{w}_\gamma^i\}_{i=1}^L$ и $\{\sqrt{\eta_i}\}_{i=1}^L$ се счита, че алгоритъмът е обучен, като за изчисляването на теглата $\{\mathbf{w}_\gamma^i\}_{i=1}^L$ се използва представянето за паралелни изчисления (3.18). Във фазата Тест, за данни представени чрез вектора \mathbf{x} , първо се генерира представянето със случайни признаци (3.16), след което се изчислява условната вероятност (3.21) за всеки един от класовете и този вектор се причислява към класа, за който условната вероятност е най-голяма. Това е изобразено на Фиг. (3.5).

3.4 Основни резултати и изводи

Основните резултати от работата в тази глава са следните:

1. Разработен е алгоритъм за разпознаване на изражения на лица с локални признаци и изборът им, включващ:
 - Предварителна обработка, осигуряваща детектиране на лице в естествена сцена и регистрация в обща координатна система.
 - Извличане на log-Gabor признаци, които са устойчиви на вариации в осветлението, използвайки конволюция в честотна област.
 - Избор на признаци чрез разрежена регресия, която запазва вграждането на многообразието, моделиращо геометрията на данните и дискриминативността.
 - Класификатор, реализиран с SVM с ядро.
2. Разработен е алгоритъм за разпознаване на изражения на лица за големи обеми от данни, включващ:
 - Предварителна обработка, осигуряваща детектиране на лице в естествена сцена.
 - Извличане на геометрични признаци на лицето, представени чрез ключови точки, чрез алгоритъм с обучение и високо бързодействие.
 - Класификация на израженията чрез мултономиална регресия, реализирана с множество ядра, като всяко ядро е апроксимирано чрез случайни признаци базирани на хармонични функции.

Изводите от работата по тази глава са следните:

1. Признаци извлечени с оператора log-Gabor са подходящ избор при разпознаването на изражения на лица.

2. Изборът на признаци е необходим за случаите, когато извлечените признаци са с по-голяма размерност от размерността на изображението.
3. Геометричните признаци на лицето представени чрез ключови точки предоставят основната информация при разпознаването на изражения.
4. Апроксимацията на метода с ядрото чрез случайни признаци е надежден подход, предоставящ ефективност, сходна със съвременните невронни мрежи.

3.5 Приноси към Глава 3

Научните приноси са:

1. Разработен е комбиниран метод, позволяващ работа с големи обеми от данни, базиран на геометрични признаци на лицето и обучение чрез множество ядра, апроксимирани със случайна извадка от косинусови базиси. Този принос е към Глава 3, точка 3.3.

Научно-приложните приноси са:

1. Разработени са два алгоритъма за разпознаване на изражения на лица, базирани на локални и геометрични признаци, последващ избор на признаци чрез разредена регресия и апроксимация на метода с ядрото със случайна извадка от косинусови базиси. Избора на признаци и апроксимацията на метода с ядрото подобряват дискриминативността и позволяват работа с големи обеми от данни. Този принос е към Глава 3, точки 3.2 и 3.3.

Приложните приноси са:

1. Програмна реализация и експериментални изследвания на алгоритъм за разпознаване на изражения на лица, използващ разреден избор на log-Gabor признаци. Постигнат коефициент на разпознаване от алгоритъма 96.7%, подобряващ докладваните в литературата резултати, използвайки стандартна база данни от лица с анотирани изражения. Този принос е към Глава 3 и 5, точки 3.2 и 5.2.1.
2. Програмна реализация и експериментални изследвания на алгоритъм за разпознаване на изражения на лица, използващ извличане на геометрични признаци на лице и апроксимация на метода с ядрото. Постигнат коефициент на разпознаване от алгоритъма 53.6%, подобряващ докладваните в литературата, използвайки стандартна база данни от лица с анотирани изражения. Този принос е към Глави 3 и 5, точки 3.3 и 5.2.2.

Алгоритмите за разпознаване на изражения на лица, разработени в тази глава, са докладвани в следните научни публикации:

- [D2] **K. Tonchev**, N. Neshov, A. Manolova, V. Poulkov, "Expression recognition using sparse selection of log-Gabor facial features. In 2017 *Fourth International Conference on Mathematics and Computers in Sciences and in Industry (MCSI)*, (pp. 11-15). IEEE.

- [D3] **K. Tonchev**, N. Neshov, A. Manolova, T. Sechkova, "Applications of large scale kernel machines for real world human mood estimation." *International Journal of Signal Processing, Volume*, vol. 2, 2017, ISSN: 2367-8984

4 Разработване на метод и алгоритми за разпознаване на 3D изображения на лица

Едно от най-важните предимства на разпознаването на лица, пред другите биометрични технологии за идентификация е, че методът е неинванзивен. Това означава, че за идентификацията не е нужно сътрудничеството на наблюдаваният индивид, нито предприемането на конкретни действия от негова страна. Това предимство също така се явява и потенциален недостатък. Например, човек може да претендира за различна идентичност, просто чрез въвеждане в системата за разпознаване на снимка на лицето, което той претендира да бъде. Поради тази причина, в последните години, вниманието на изследователите е насочено към разработване и изследване на методи за разпознаване на 3D лица. Смята се, че те са по-подходящи за ускорена персонална идентификация във видеоследователности и в случай на неконтролирани условия на заснемане [148]. Разработените през последните години нови и по-евтини технологии за 3D сканиране на лицеви структури, решават до голяма степен втория основен проблем на разпознаването на 3D лица, а именно, лесно и евтино създаване на подходящи бази данни от 3D изображения на лица [149]. Повечето разработки за разпознаване на 3D лица се занимават с вариациите в израженията на лицето, но не се фокусират върху устойчивостта при деградация на качеството на 3D моделите на лицата. Тези деградации са, например, осветеност, движение пред 3D сензора, липсващи данни поради припокриване или погълната светлина или голямо разстояние до 3D скенера [150]. Освен това, компресията и дискретизацията (ограничаване на разделителната способност на 3D модела на лицето), също могат да доведат до деградиране на моделите. Алгоритмите докладвани в литературата, използват, както локални, така и холистични признаци при разпознаването на лицето. За разлика от 2D изображенията, при 3D изображенията се предоставя геометрията на лицето, което предполага и използването на друг клас алгоритми извличащи геометрични признаци, експлицитно или имплицитно.

В текущата глава се предлагат два алгоритъма за разпознаване на 3D лица. Първият алгоритъм представя 3D модела на лицето като дълбочинно изображение, което има аналог с 2D изображенията на лица, защото разстоянието между скенера и лицето е дискретизирано върху равномерна двумерна решетка. Използвайки това представяне, се извличат холистични и локални признаци, последвани от намаляване на размерността и класификатор, за да се разпознае лицето. Вторият алгоритъм използва представянето на лицето като облак от 3D точки. Извличат се локални признаци представени чрез нормали към повърхнина и адаптивни локални регресионни ядра. Локалните признаци се комбинират използвайки множество ядра, което се използва в SVM класификатор за разпознаването на лицето.

4.1 Анализ на литературните данни по разпознаване на 3D изображения на лица

Представянето на лицето и процесът на неговата идентификация, е активна изследователска област през последните години, с основни акценти като определяне на позата, откриване и разпознаване на лица в, неподвижни изображения и видео. Използването на 3D лицеви изображения за подобряване ефективността на системите за разпознаване на лица е идентифицирано като средство, което превъзхожда разпознаването на лица от 2D изображения. В действителност, разпознаването на 3D лица е по-благоприятно, защото е по-малко чувствително към вариации в осветеността и е инвариантно към деформации на лицето, предизвикани от промени в изражението. Методите за разпознаване на 3D лица могат да бъдат групирани спрямо различни критерии [151], като например класически или съвременни, холистични или локални, единична модалност или много-модални и т.н. Най-често срещаното групиране се прави между холистичните и локални методи, защото се оказва, че локалните методи са по ефективни спрямо холистичните [152].

Холистичните методи обработват цялата лицева повърхност, което обикновено ги прави компактни и следователно, с по-бързи. Те са относително стабилни при наличие на шум, но са чувствителни към нормализацията на изображението на лицето, както и закриването му. Точността на разпознаването намалява, ако в 3D модела на лицето са включени елементи като косата, ушите и врата [153]. За разлика от холистичните, локалните методи делят лицето на прозорци (региони) на интерес и за всеки от тях извличат признаци. В някои случаи тези прозорци могат да бъдат намалени до много малки региони около ключови зони, като очите. Характерна черта на локалните методи е техния потенциал да се справят с припокривания и изражения, като това основно се дължи на комбинацията от много локални дескриптори [154]. Много от най-новите подходи за 3D разпознаване на лица са използвали локални признаци, за които са докладвани високи степени на разпознаване, използвайки някои еталонни бази данни [155].

Друго сравнение, което е важно за разпознаването на 3D лица, е групиране на алгоритмите по устойчивост спрямо конкретен тип вариация. Така, групите могат да се разделят на устойчиви спрямо поза, изражения и покривания. Алгоритъм за разпознаване на 3D изображения на лица, устойчив на изражения, е предложен в работата [156]. В основата на тази работа, стои извличането на признаци, наименувани Angular Radial Signature (ARS), които се извличат от региони на лицето считани се като полу-твърди, като например бузите. На всеки един от признаците ARS се намалява размерността с PCA, реализиран с ядро. Всички признакови вектори с намалена размерност се комбинират в един вектор, който се подава на класификатора SVM за разпознаване на лицето. Предложеният от тях алгоритъм е тестван на базите данни FGRC v2.0 [157] и SHREC 2008 [158], като докладват висока ефективност.

В работата [159] е предложено разпознаването на лицето да се извърши използвайки представянето с мрежа (mesh). Първоначално се детектират ключови точки с минималната и максимални кривини на повърхнината на лицето. След което, в околност на тези точки, се извличат микро-компоненти на повърхнината и се изчислява сходство между отделни лица, с което се извършва класификацията. Алгоритъмът е тестван използвайки базата данни Bosphorus [160], и на база

на резултатите се доказва, че предложеният алгоритъм е устойчив при промяна на изражението на лицето.

Представянето на лицето с мрежа (mesh), за разпознаване на 3D лица е предложено в [161]. Върху това представяне се детектират ключови точки, използвайки дескрипторите Difference of Gaussians (DoG) и Gromov-Hausdorff (GH), за които е предложено оригинално решение с цел подобряване стабилността при изчисляването на ключовите точки и дескрипторите. Предложеният алгоритъм е тестван с базата данни Bosphorus и е докладвана висока ефективността на алгоритъма.

В работата [162] се предлага разделяне на лицето на серии от ленти, които се намират на равни геодезични разстояния. След това се построява таблица, която описва индекса на посоката на всяка една такава лена, измервайки пространственото отместване между две ленти. Две лица се сравняват чрез индексите на таблиците им, с което се постига ниска изчислителна сложност, което е предимство в случаите на големи бази данни. За тест на алгоритъма се използва базата данни SHREC 2008 и се докладва висока ефективност на алгоритъма, спрямо други използващи същата база данни.

Алгоритъм за разпознаване на 3D лица, който е устойчив на покривания на лицето, е предложен в [163]. Използва се ротация на облак от точки, множество проекции, и сравняване на ключови точки чрез гласуване. Основната идея е да се проектира облака от точки върху 2.5D (дълбочинно изображение) чрез прилагане на ротация. След това се извличат ключови точки, които са предварително определени. Алгоритъмът е тестван върху базата данни GavabDB [164] и е докладвана точност от 95% на разпознаване при неутрални лица, и 90% на разпознаване при лица с различни изражения.

В работата [165] се предлага нов алгоритъм, който е устойчив на покривания на лицето от коса пред лицето, ръце и очила. Първоначално се извършва регистрация на лицето, състояща се от две стъпки. Първата стъпка региона на носа, използвайки карта на изпъкналите повърхнини, след което използва алгоритъма Iterative Closest Point (ICP) [166], за да регистрира лицето. Регионът, в който се намира закриването се открива, като се използва обобщен модел на лицето. След това се премахва този регион който не принадлежи на лицето и тя се възстановява използвайки PCA. Използвана е базата данни Bosphorus за тест на алгоритъма и е докладвано подобряване на резултатите от 76.12% до 94.23%.

В работата [167] е представен локален подход към 3D разпознаване на лице в комбинация с функция за избор на модел за изследване на относителното значение на различните региони на лицето, за целите на търсене на различие между отделните раси. Предложеното решение е тествано с помощта на база данни FRGC. Резултатите от експериментите доказват количествено предположението, че различните региони на лицето имат различно значение за нуждите на разпознаването, като показват и съществени различия на лицевите региони за различните етнически групи.

4.1.1 Изводи от литературния анализ по разпознаване на 3D изображения на лица

Изводите от анализа на литературата по определяне на възраст от изображения на лица могат да се обобщят до следните:

- Холистичният подход при разпознаването на 3D изображения на лица изисква предварителна регистрация, която сама по себе си трябва да е с много голяма точност. Също така, този подход не е устойчив при наличието на закриване на лицето, като очила или поставена ръце.
- Независимо от подхода е необходимо да се извърши предварителна филтрация, за да се отстрани точков шум или попълнят липсващи данни, които възникват при процеса на сканиране.
- Сканираните 3D изображения на лица са устойчиви по отношение на смяна на осветлението, освен това са по-лесни за регистрация, защото са представени в метрични мерни единици, което ги прави подходящи за прилагането на локални признаци които вземат предвид геометрията в прозореца на интерес.
- Локалните методи подчертано работят по-добре при разпознаването на 3D изображения на лица, предоставяйки устойчивост на вариации в позата и закриване.

4.2 Разпознаване на 3D лица чрез дълбочинно изображение и линейни статистически трансформации

За да се постигне висока точност на сканирането на геометричната форма на лицето, то се извършва с лазерен скенер. Резултатът от сканирането е облак от точки в 3D пространството, като борят и гъстотата им зависи от точността на скенера [150]. За разлика от 2D изображенията, недостатък на облака от точки е неговото не-равномерно отчитане на измерванията в пространството. Макар и да може в тази си форма да се използва за обработка, често се налага форматиране на облака като мрежа (mesh) или дълбочинно изображение. Представянето с мрежа (mesh) съдържа информация за свързаност между отделните точки от облака, докато дълбочинното изображение, още наименувано 2.5D, форматира облака така, че хоризонталната и вертикална координати да са равномерно дискретизирани, а третата координата да представлява разстоянието от скенера до лицето, тоест дълбочината. Това представяне е удобно за работа, защото наподобява двумерните изображения, и съответно могат да се прилагат всички алгоритми, които са разработени за тях.

В тази точка е представен ново-предложен алгоритъм за разпознаване на 3D изображения на лица, който използва дълбочинни изображения и извлича признаци, които характеризират геометрията на лицето в локална област. Самите признаци са инвариантни по отношение на ротация и това, че са локални, означава, че алгоритъмът е устойчив при изражения на лицето, както беше установено в литературното проучване в т. 4.1. За да се намали размерността на признаците се използва PCA, последван от LDA, с цел увеличаване на дискриминативността и допълнително намаляване на размерността. Така извлечените признаци се подават на класификатора SVM, реализиран с ядро. Предимството на предложенният алгоритъм, спрямо другите предложени в литературата, е *извличането на*

геометрични характеристики на лицето, но представено като дълбочинно изображение.

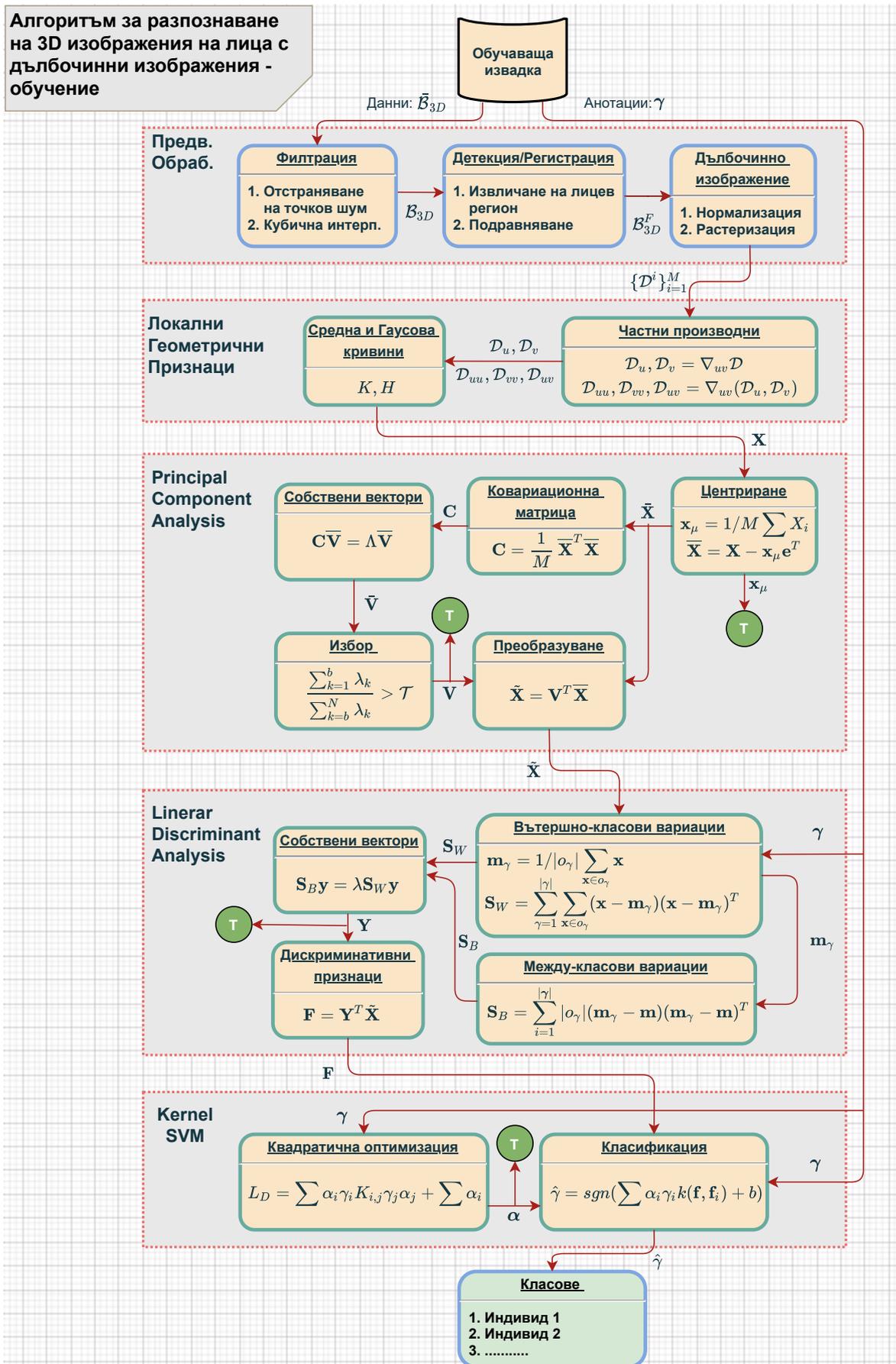
4.2.1 Блок-схема на потока на информацията и анализ на предложения алгоритъм

Разпознаването на изображения на лица се извършва в две фази, Обучение и Тест, и следва общият подход за разпознаване на изображения на обекти, описан в т. 1.1.2. Извличането на признаци се извършва от няколко последователни оператора, като коефициентите на първият оператор са фиксирани, но коефициентите на всеки един от последващите оператори се изчисляват през фазата Обучение. Блок-схема на потока на информацията във фазата Обучение, е представен на Фиг. 4.1.

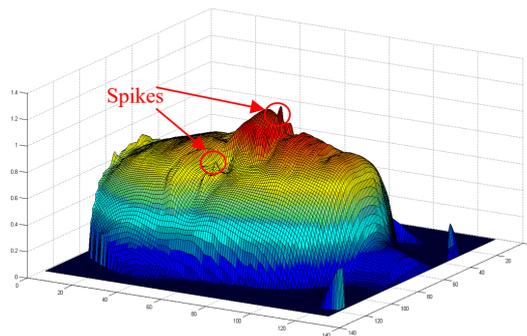
На алгоритъма се подават 3D изображения $\bar{\mathcal{B}}_{3D}$, които са сканирани области от лицето, но съдържат и области около лицето, включващи врат, коса, уши и също така шум поради недостатъци на процеса на сканиране. Това обуславя нуждата от предварителна обработка, която да подготви лицевата област за разпознаване.

В първата стъпка от фазата на обучение на предложения алгоритъм, **Предварителна Обработка**, се извършва подготовка на 3D изображенията на лицата за извличане на признаци. В блока *Филтрация* се извършват две операции. Първата е *Отстраняване на точков шум*, при която се използва медианен филтър за да се филтрират „шипове”, които са грешки при сканирането, свързани най-вече с отражения от огледални повърхности или грешки на скенера. Лице с такъв тип шум е визуализирано на Фиг. 4.2, като за по-добра визуализация на фигурата е изобразена версията с мрежа (mesh).

Медианният филтър се прилага, като за всяка околност се вземат 9-те най-близки съседа и се подредят по нарастване на z координатата, след което се вземе централната, 5-та точка. След като се филтрира изображението се прилага *Кубична интерполация* за да се запълнят липсващи точки или области. Такива ситуации често възникват когато лъчът от лазера не е отразен, например в областите около клепачите на очите или устата. Така подготвени, 3D изображенията \mathcal{B}_{3D} се подават на етапа *Детекция/Регистрация*, в първата стъпка на който се извлича лицевият регион. Алгоритъмът за извличане се състои в намиране на върха на носа и позициониране на сфера с предварително зададен радиус в тази точка. Извлеченото лице представляват точките, които са вътре в тази сфера. Върхът на носа се детектира като точката, най-близо до скенера, тоест с най-малка стойност z . Това е валидна операция, защото е отстранен точковият шум. Регистрацията се състои в подравняване при което върхът на носа се избира да бъде началото на координатната система, тоест извършва се трансляция. Като резултат, лицевите региони \mathcal{B}_{3D}^F съдържат облак от точки, който е филтриран и съдържа само лицевата част от 3D изображенията. Те се подават на етапа *Дълбочинно изображение*, в който облакът от точки на лицевата област се преобразува в дълбочинно изображение. Това става като се извърши нормализация по z остта и растеризация, която причислява всяка една точка от облака към двумерна равномерна решетка. Растеризацията се извършва чрез нормализация по x и y координатите на облака от точки, последвано от линейна интерполация. Възможно е използването и на друг вид интерполация, но сканираните изображения със съвременни скенери са с



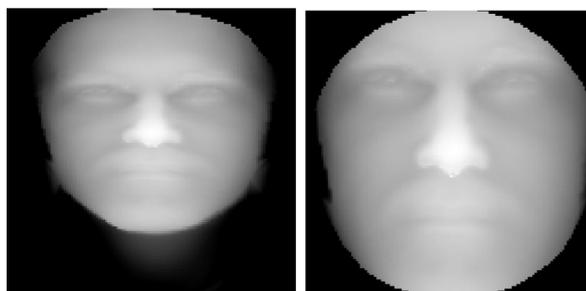
ФИГУРА 4.1: Блок-схема на алгоритъм за разпознаване на 3D изображения на лица, представени с дълбочинно изображение - фаза Обучение.



ФИГУРА 4.2: Визуализация на точков шум на 3D изображение на лице.

достатъчно висока разделителна способност, което отхвърля нуждата от по-точна интерполация. Нормализацията преобразува динамичният диапазон на лицето в интервала 0-1, като по този начин изравнява динамичният диапазон на всички лица, което е сходно на хистограмна еквиализация на изображенията. Обработените лицеви изображения се подреждат в множеството $\{\mathcal{D}^i\}_{i=1}^M$ от дълбочинни изображения, което се подава за извличане на признаци. Примерно 3D изображение и дълбочинното изображение, което е извлечено от него, са представени на Фиг. 4.3. Използваните стъпки от предварителната обработка са мотивирани от анализа на литературата направен в предходната точка, и удовлетворяват следните условия:

- Медианната филтрация и интерполацията премахват точков шум и погълват липсващи точки, които са често срещани проблеми заради недостатъците на лазерните скенери.
- Изрязва се лицевият регион и се премахват ненужни области като коса и врат. Лицевите региони се регистрират в обща координатна система, с което се намаляват вариациите по отношение на транслация.
- Изравняват се динамичните диапазони на всички лица и се растеризират, с което се намаляват вариациите свързани с геометрията, и се подготвят за извличане на признаци.



ФИГУРА 4.3: Сканирано 3D изображение и извлеченото дълбочинно изображение.

Следващата стъпка от обработката е извличането на **Локални Геометрични Признаци**. Тези признаци се базират на кривината и по-точно на Средната и

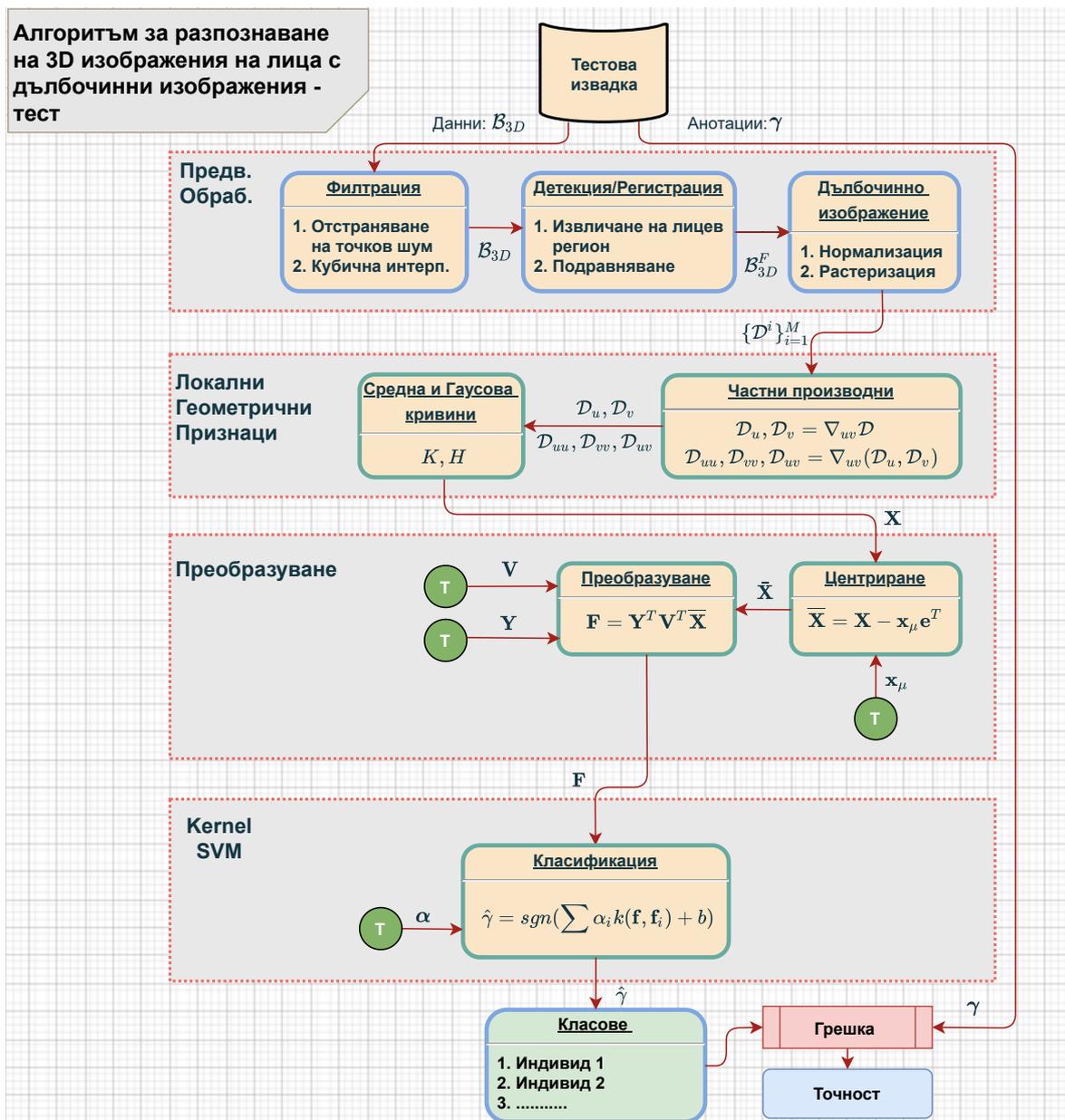
Гаусови кривини K, H . При изчисляването на тези кривини се използват първите и вторите частни производни $\mathcal{D}_u, \mathcal{D}_v, \mathcal{D}_{uu}, \mathcal{D}_{vv}, \mathcal{D}_{uv}$, изчислени по двете координати на изображението u, v . Мотивацията за избор на този тип признаци е следната:

- Те са локални, и както беше установено в литературния анализ, локалните признаци са за предпочитане, защото извличат информация която е устойчива по отношение на лицеви изражения.
- Тези признаци са инвариантни по отношение на ротация и трансляция [168], което ги прави устойчиви по отношение на грешки в регистрацията.
- Описват геометрични характеристики на повърхнината на лицето.

Извлечените признаци са с по-висока размерност, поради което тя се намалява с алгоритъма PCA, за да се избегне ефекта от „проклятието на размерността“. След това, върху данните с намалена размерност $\tilde{\mathbf{X}}$ се прилага етапа **Linear Discriminant Analysis**, реализиран с алгоритъма LDA. Първоначално се изчисляват *Вътрешно-класови вариации*, които се кодират в матрицата \mathbf{S}_W , а след тях се изчисляват *Междо-класови вариации*, които се кодират в матрицата \mathbf{S}_B . С тези матрици се изчисляват *Собствени вектори*, чрез задачата за обобщените собствени стойности и собствени вектори. Собствените вектори се съхраняват в матрицата \mathbf{Y} , която се запазва за фазата Тест, както е показано на фигурата с Т. Броят на собствените вектори е $|\gamma| - 1$, тоест броят на класовете минус едно, и с тях се извършва извличането на *Дискриминативни признаци*, чрез линейното преобразуване $\mathbf{F} = \mathbf{Y}^T \tilde{\mathbf{X}}$. Прилагането на PCA, и последващото прилагане на LDA, е мотивирано от нуждата за намаляване на размерността и подчертаване на дискриминативността. Последната стъпка от предложеният алгоритъм е прилагането на алгоритъма SVM, използвайки \mathbf{F} , за да се определи класовата принадлежност $\hat{\gamma}$ на всяко едно от изображенията на обучаващата извадка.

Фазата тест се прилага, използвайки тестова извадка, за която са известни анотациите γ . Блок-схема на тази фаза е представена на Фиг. 4.4. Предварителната обработка и извличането на геометрични признаци в тази фаза са идентични с тези на фазата Обучение. Стъпката **Преобразуване** цели извличането на дискриминативни признаци като използва линейните преобразувания от фазата на обучение (на PCA и LDA) \mathbf{Y}, \mathbf{V} . Тяхно прилагане се състои в умножение на матрица по вектор, което допуска те да бъдат умножени предварително, с което се спестяват изчисления. Последната стъпка е класификация с класификатора SVM реализиран с ядро, който използва опорните вектор α от фазата Обучение за да определи класа (идентитета) на изображението $\hat{\gamma}$. Ако определеният клас съвпада с анотацията, то грешката се счита за 1, ако не съвпада, грешката се счита за 0 (блок *Грешка*). Акумулирането на всички грешки и разделянето им на броя на изображенията в тестовата извадка задава точността на алгоритъма, което е и неговата ефективност (блок *Точност*). Понякога, ефективността се задава в проценти като се умножи по 100.

Подробно описание на етапите от алгоритъма е представено в следващите точки.



ФИГУРА 4.4: Блок-схема на алгоритъм за разпознаване на 3D изображения на лица описваща фазата Тест.

4.2.2 Формиране на дълбочинно изображение

Като резултат от регистрацията, лицевите региони са извлечени и подравнени в обща координатна система. Първата стъпка при преминаването към дълбочинното изображение на лице е намирането на минималните и максималните стойности на облака от точки $\{x_i, y_i, z_i\}_{i=1}^N$:

$$\begin{aligned}x_{min} &= \min(x_i) & x_{max} &= \max(x_i) \\y_{min} &= \min(y_i) & y_{max} &= \max(y_i) \\z_{min} &= \min(z_i) & z_{max} &= \max(z_i)\end{aligned}\tag{4.1}$$

за всяка точка от облака $i, i = 1, \dots, N$. След това се построява двумерна равномерната решетка с която се извършва растеризация като първоначално, изображението се транслира по x и y като се извадят минималните стойности съответно по двете координатни оси. След това, използвайки x_{max} и y_{max} , координатите се нормализират като се запазва съотношението по хоризонтала и вертикала. Като крайна стъпка от растеризацията се прилага линейна интерполация, която да трансформира нормализираните координати по x и y на облака от точки към предварително дефинираната двумерна равномерна решетка. Тази равномерна двумерна решетка представлява координатната система на дълбочинното изображение.

След като е растеризирано, дълбочинното изображение, то се нормализира в диапазона 0-1, като по този начин изравнява динамичният диапазон на всички лица от обучаващата и тестова извадки.

4.2.3 Извличане на признаци от дълбочинно изображение чрез кривината на формата

Кривината предоставя информация за характера на повърхнината в околност на точка, което я прави подходящ инструмент при анализа на тримерни изображения на лица [168]. Изчисляването на кривината за повърхнини е доста по-сложно спрямо това за криви, тъй като повърхнините се изкривяват не-изотропно. Използвайки кривината на кривите, които лежат на повърхнините обаче, може да се постигне характеризиране на кривините на повърхнините. Изчислена по този начин кривината на повърхнина е неточно, но е достатъчно, за да бъде характеризирана повърхнината напълно. Това приближение предоставя също и подходяща мярка с която може да се оцени достатъчно точно приближението на повърхнината.

В предложения алгоритъм за извличане на признаци се използват средната H и гаусова K кривини, защото са инвариантни към ротация и трансляция [168]. За изчисляването им се използват частните производни по независимите координати на изображението (u, v) , [169]. Нека е дадено 3D изображение на лице \mathcal{D} , тогава

първите и вторите частните производни по независимите координати се изчисляват, чрез приближение от първи ред:

$$\begin{aligned}
 \mathcal{D}_u &= \mathcal{D}(u, v) - \mathcal{D}(u + 1, v) \\
 \mathcal{D}_v &= \mathcal{D}(u, v) - \mathcal{D}(u, v + 1) \\
 \mathcal{D}_{uu} &= \mathcal{D}_u(u, v) - \mathcal{D}_u(u + 1, v) \\
 \mathcal{D}_{vv} &= \mathcal{D}_v(u, v) - \mathcal{D}_v(u, v + 1) \\
 \mathcal{D}_{uv} &= \mathcal{D}_u(u, v) - \mathcal{D}_u(u, v + 1).
 \end{aligned} \tag{4.2}$$

Приближението от първи ред е неточно при наличието на шум, но в предложеният алгоритъм то се използва, защото в предварителната обработка изображението е филтрирано с медианен филтър, който не само премахва точковия шум, но и гаусовия. Използвайки частните производни, средната и гаусови кривини се изчисляват както следва [169]:

$$\begin{aligned}
 K &= \frac{\mathcal{D}_{uu}\mathcal{D}_{vv} - \mathcal{D}_{uv}^2}{(1 + \mathcal{D}_u^2 + \mathcal{D}_v^2)^2} \\
 H &= \frac{(1 + \mathcal{D}_u^2)\mathcal{D}_{vv} + (1 + \mathcal{D}_v^2)\mathcal{D}_{uu} - 2\mathcal{D}_u\mathcal{D}_v\mathcal{D}_{uv}}{2(1 + \mathcal{D}_u^2 + \mathcal{D}_v^2)^{3/2}}.
 \end{aligned} \tag{4.3}$$

Изчислените кривни H и K са реални числа, което означава, че на един елемент от дълбочинното изображение се генерират два нови, с което се увеличава размерността два пъти.

4.2.4 Извличане на дискриминативни признаци с Linear Discriminant Analysis

Извлечените признаци представени чрез средната и гаусови кривни повишават размерността на изображението поради което е необходимо да се приложи намаляване на размерността. Първоначално се прилага PCA, както е показано на Фиг. 4.1, с което се цели да се отстранят вариации като шум. След това се прилага LDA, с което допълнително се намаля размерността и се цели постигане на дискриминативност. Математическите означения в тази точка са идентични с тези на Фиг. 4.1, и фигурата може да се използва като референция.

При LDA алгоритъма се търсят $|\gamma|-1$ направления от типа $f_i = \mathbf{y}_i^T \tilde{\mathbf{x}}$, максимизиращи дискриминативността на данните [38], където $\tilde{\mathbf{x}}$ е признаковият вектор, резултат от PCA алгоритъма. Първа стъпка от LDA алгоритъма е изчисляването на вариациите, използвайки ковариационните матрици на ниво клас:

$$\begin{aligned}
 \mathbf{S}_W &= \sum_{c=1}^{|\gamma|} \sum_{\mathbf{x} \in o_\gamma} (\mathbf{x} - \mathbf{m}_\gamma)(\mathbf{x} - \mathbf{m}_\gamma)^T, \\
 \mathbf{S}_B &= \sum_{i=1}^{|\gamma|} |o_\gamma| (\mathbf{m}_\gamma - \mathbf{m})(\mathbf{m}_\gamma - \mathbf{m})^T,
 \end{aligned} \tag{4.4}$$

където $\mathbf{m}_\gamma = 1/|o_\gamma| \sum_{\mathbf{x} \in o_\gamma} \mathbf{x}$ е средната стойност на данните за всеки клас, а \mathbf{m} е средната стойност на всички данни. За да се постигне дискриминативност на

проекциите, те трябва да са такива, че да се минимизират вътрешно-класовите вариации, докато съвременно се максимизират между-класовите вариации. Това се постига като се решава частното на Релей:

$$\arg \max_{\mathbf{Y}} \frac{\mathbf{Y}^T \mathbf{S}_B \mathbf{Y}}{\mathbf{Y}^T \mathbf{S}_W \mathbf{Y}}, \quad (4.5)$$

като решенията на това частно се задават с решенията на обобщената задача за собствени стойности:

$$\mathbf{S}_B \mathbf{y} = \lambda \mathbf{S}_W \mathbf{y}. \quad (4.6)$$

Матрицата \mathbf{S}_B е сума от $|\gamma|$ матрици с ранг едно, където най-много $|\gamma|-1$ матрици могат да бъдат линейно независими. Следователно съществуват най-много $|\gamma|-1$ ненулеви собствени стойности, които представляват дискриминативните проекции. Собствените вектори, съответстващи на ненулевите собствени вектори се нареждат в матрицата \mathbf{Y} и извличането на дискриминативни признаци се извършва чрез проекцията в подпространство:

$$\mathbf{F} = \mathbf{Y}^T \tilde{\mathbf{X}}, \quad (4.7)$$

където \mathbf{F} са извлечените признаци с размерност $\mathbf{F} \in \mathbb{R}^{M \times (|\gamma|-1)}$.

4.3 Разпознаване на 3D лица с комбиниране на допълващи се локални признаци чрез множество ядра

Използването на много и разнообразни признаци, в задачите за разпознаване на изображения на обекти, е интуитивен подход, защото така се гарантира извличането на колкото се може повече информация от изображението, която е свързана със задачата. Проблемът който възниква в тази ситуация е как да бъдат комбинирани признаците, така че да се максимизира ефективността на алгоритъма. В предложението в тази точка алгоритъм се извличат два вида признаци, адресирайки различни характеристики на повърхнината на 3D изображението на лицето. Тези признаци се комбинират чрез множество ядра и се използват като резултантно ядро в класификатора SVM. Комбинирането на ядрата става чрез изпъкнала комбинация от ядра, чиито коефициенти се изчисляват през фазата Обучение. Нещо повече, цели се изпъкналата комбинация да бъде разреждана с което може да се каже, че се прави както избор на признаци, така и на оптималните параметри на ядрата. Предимството на предложеният алгоритъм, спрямо другите предложени в литературата, е *използването на допълващи се геометрични признаци, и тяхната комбинация чрез множество ядра, като резултантното ядро се използва в класификатора SVM.*

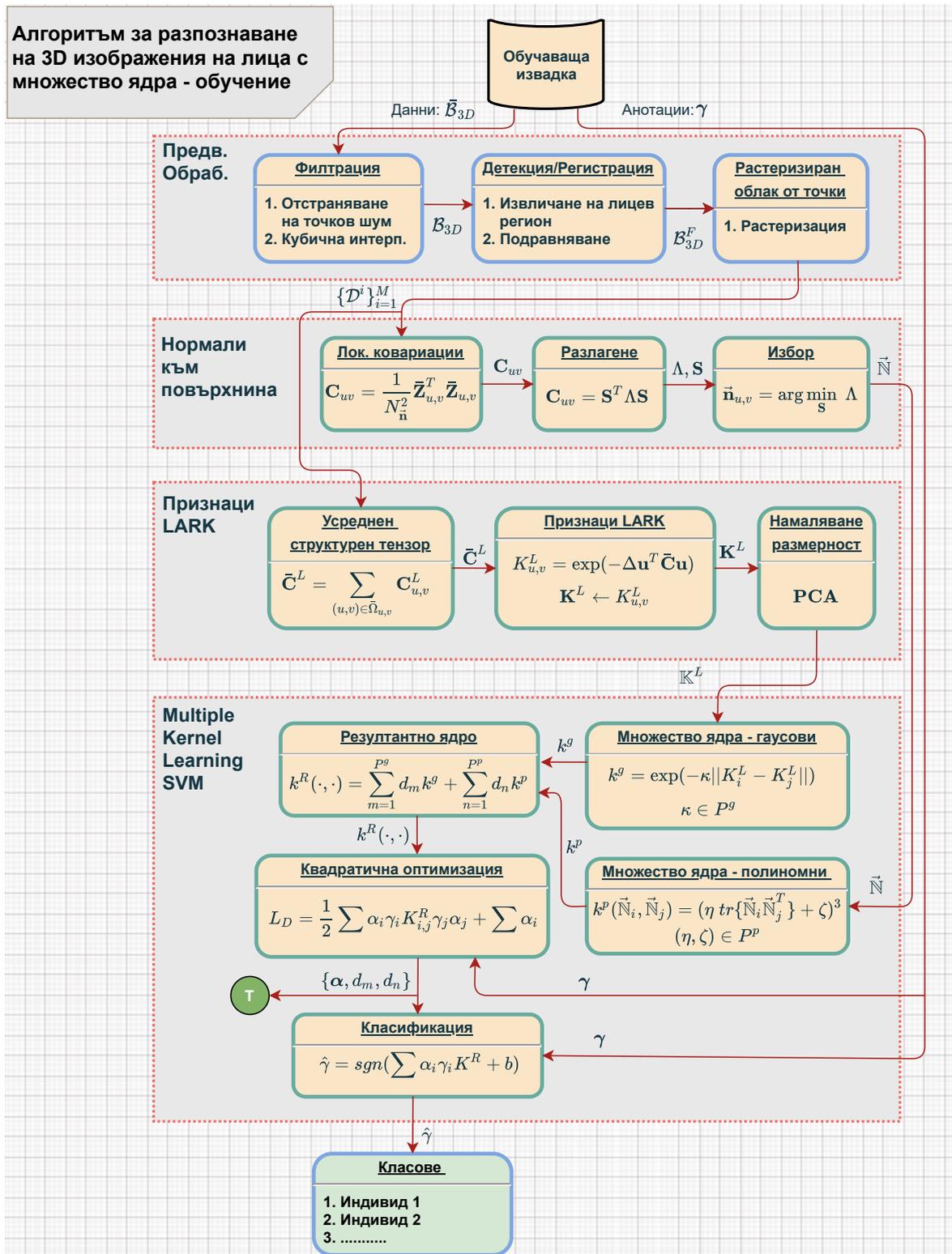
4.3.1 Блок-схема на потока на информацията и анализ на предложения алгоритъм

В предложения алгоритъм разпознаването на 3D изображенията на лица се извършва в две фази, Обучение и Тест, като следва общата рамка описана в т. 1.1.2. Извличането на признаци се извършва с няколко последователни оператора, на които коефициентите не зависят от обучаващата извадка, а само от локалната статистическа информация на изображението. Извлечените коефициенти се използват за обучение на SVM класификатор, реализиран с множество ядра, коефициентите на които, също се изчисляват по време на обучението. Блок-схема на потока на информацията във фазата Обучение е представен на Фиг. 4.5.

В първата стъпка от фазата Обучение на предложения алгоритъм, **Предварителна Обработка**, се извършва подготовка на 3D изображенията на лицата за извличане на признаци. В блока *Филтрация* се извършват две операции. Първата е *Отстраняване на точков шум*, при която се използва медианен филтър за да се филтрират „шипове“, които са грешки при сканирането, свързани най-вече с отражения от огледални повърхности или грешки на скенера. Втората операция е *Кубична интерполация* с която се запълват липсващи точки, след което филтрираните изображения \mathcal{B}_{3D} се подават на блока *Детекция/Регистрация*. Алгоритъмът за извличане на лицевия регион се състои в намиране на върха на носа, върху който се центрира сфера с предварително зададен радиус, и се избират точките, които са вътре в сферата. Върхът на носа се детектира като точката, най-близо до скенера, тоест с най-малка стойност z . Така лицевите региони \mathcal{B}_{3D}^F съдържат облак от точки, който е филтриран и представлява само лицевата част от 3D изображенията. Те се подават на стъпката *Растреризация на облак от точки*, при която облакът от точки на лицевата област се растреризира по x и y координатите по двумерна равномерна решетка. След растреризацията се прилага линейна интерполация за да се намалят артефактите. Така получените лицеви изображения се подреждат в множеството $\{\mathcal{D}^i\}_{i=1}^M$, което се подава за извличане на признаци. Предварителната обработка в предложеният алгоритъм е същата като тази, описана в т. 4.2.1 с тази разлика, че не се нормализира динамичния диапазон по z , с което се запазва изображението като облак от точки, но с равномерна решетка по x и y , което улеснява последващите обработки.

В предложения алгоритъм се извличат два вида признаци независимо един от друг. Първият тип са нормалните към всяка една точка от изображението $\vec{n}_{u,v}$, те се извличат в етапа **Нормали към повърхнината**. Изчисляването се извършва, като се разлага локалната ковариационна матрица \mathbf{C}_{uv} по собствени стойности, и се избере собственият вектор, който съответства на най-малката собствена стойност $\vec{n}_{u,v}$, както е показано Фиг. 4.5. Мотивацията за избор на нормалите е, че те закодират скоростта на промяна на повърхнината в околност на точка [170], което ги прави инвариантни към ротация и трансляция. Също така, те са с норма единица, което ги прави и подходящи при интерпретация и изчисления.

Вторият тип признаци са наименувани Locally Adaptive Regression Kernels ([LARK](#)) и са предложени в [171]. Тези признаци представляват претеглена апроксимация на локалните геодезични разстояния в околност на точка. Изображението се обработва на прозорци, центрирани във всеки елемент от него. Геодезичното разстояние се изчислява като се използва локалният структурен тензор \mathbf{C}^L , изчислен



Фигура 4.5: Блок-схема на алгоритъм за класификация на 3D изображения на лица с множество ядра - фаза Обучение.

използвайки елементите в прозореца. Както е показано на Фиг. 4.5, структурният тензор не се използва сам по себе си, а усреднена версия на него \bar{C}^L , с която се намалява влиянието на шума в първите производни. Усредняването се осъществява като се сумират няколко структурни тензора поелементно в област, която е по малка от прозореца на интерес. След усредняването се прилага и регуляризиране чрез претегляне на собствените стойности на усреднения структурен тензор (не е показано на фигурата). Използвайки усреднените и регуляризирани тензори се изчислява геодезичното разстояние между централния и околните му елементи от прозореца, след което това разстояние се претегля с отрицателна експонента в стъпката *Признаци LARK*. Така например, ако прозорецът на интерес е с размер 5×5 , за всеки елемент от изображението ще се генерират 25 нови елемента, което е увеличаване на размерността. Поради това, за да се намали размерността, се прилага алгоритъма PCA на всяко изображение независимо от останалите. Признаците LARK с намалена размерност \mathbb{K}^L се подават за обработка от класификатора. Мотивацията на избора на признаците LARK в предложения алгоритъм може да се обобщи до следното:

- Това са локални признаци които, както беше установено в изследването на литературата в началото на тази глава, са ефективни при разпознаването на 3D изображения на лица.
- Те са инвариантни по отношение на трансляция и частично ротация, защото представляват разстояние между елементи на изображението.
- Описват локални зависимости, използвайки аналитични доводи за построяване на геодезичното разстояние.
- Допълват нормалите като признаци, защото нормалите представляват диференциална характеристика.

В предложения алгоритъм за разпознаване на 3D изображения на лица се използва SVM класификатор, реализиран с множество ядра който се прилага в етапа **Multiple Kernel Learning SVM**. Множеството ядра се състои от две групи ядра, едната за нормалите, а другата за признаците LARK. Както се вижда от фигурата, за построяване на матрицата на Грам на нормалите се използва полиномно ядро от трети ред, а за LARK се използва гаусово ядро. И двете ядра са реализирани като множество ядра, генерирано с различни параметри. От множеството ядра, се генерира резултантно ядро $k^R(\cdot, \cdot)$ в стъпката *Резултантно ядро*, което представлява изпъкнала комбинация от всички ядра. Коефициентите на резултантното ядро се изчисляват по време на обучение. Обучението се извършва, като се решава дуалната формулировка на SVM. Резултатът от обучението са опорните вектори α , и коефициентите на резултантното ядро $\{d_m, d_n\}$, където d_m са коефициентите на гаусовото множество ядра, а d_n са коефициентите полиномното множество ядра. Мотивацията за избора на SVM, реализиран с множество ядра, в предложения алгоритъм, е следната:

- С множеството ядра може да се постигне комбинация от ядра, които са различни и са предназначени за различен тип данни.
- С множеството ядра може да се постигне избор на най-подходящите параметри на ядрата.

- Използва се SVM с ядро, който е оптимален класификатор.
- Съществува итеративна процедура за намиране на опорните вектори, и коефициентите на ядрата, наименувана SimpleMKL и предложена в [172].

Блок-схема на фазата Тест е представена на Фиг. 4.6. Предварителната обработка и извличането на признаците (нормалите и LARK) в тази фаза са идентични с тези на фазата Обучение. Последната стъпка е класификация с класификатора SVM реализиран с множество ядра, който използва опорните вектор α , и коефициентите на резултантното ядро $\{d_m, d_n\}$, изчислени във фазата Обучение. С тях се определя класа на изображението от тестовата извадка $\hat{\gamma}$. Ако определеният клас съвпада с анотацията, то грешката се счита за 1, ако не съвпада, грешката се счита за 0 (блок *Грешка*). Акумулирането на всички грешки и разделянето им на броя на изображенията в тестовата извадка задава точността на алгоритъма, което е и неговата ефективност (блок *Точност*). Понякога, ефективността се задава в проценти като се умножи по 100.

Подробно описание на стъпките от предложения алгоритъм е представено в следващите точки.

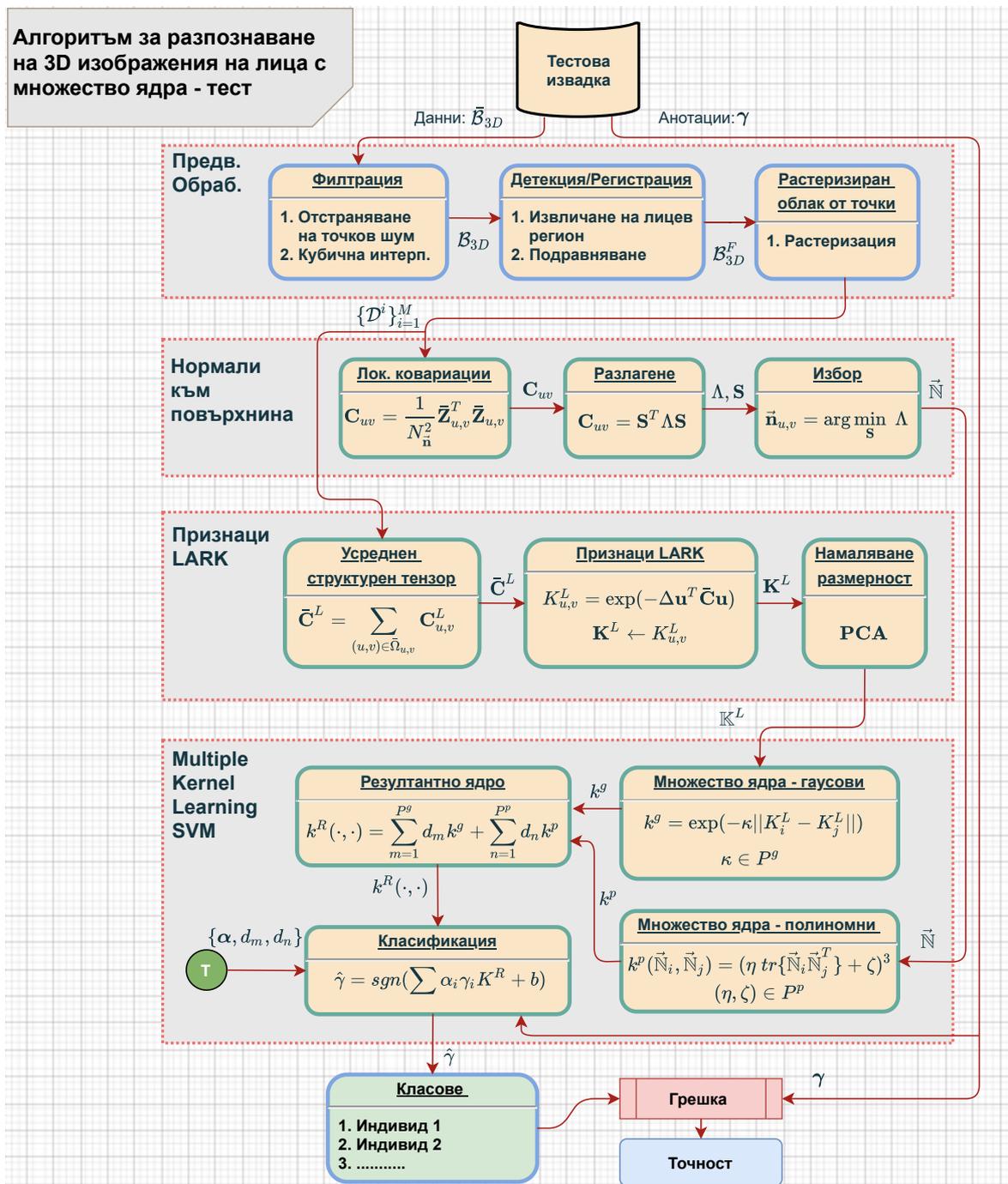
4.3.2 Извличане на признаци чрез нормали към повърхнина

Нормалите към дадена повърхнина са признаци, които са взаимствани от диференциалната геометрия и те закодират скоростта на промяна на повърхнината за локални области от нея [170]. В контекста на признаци на лицето, те представляват вектори от три компоненти сочещи навън от лицето във всяка една точка. Следователно, за всяка точка от лицевата повърхнина може да се асоциира признак, представен от нормалата в тази точка. Те могат да бъдат изчислени по различни начини, но в тази работа, те се изчисляват използвайки задачата за собствени стойности. Тъй като изображението е растеризирано по равномерна двумерна решетка, могат да се избират прозорци от него, които са с квадратни размери. Използвайки това, изчисляването на нормалите на изображението се извършва с плъзгащ прозорец, центриран върху всеки елемент от изображението. Нека този прозорец да е с размери $N_{\bar{n}} \times N_{\bar{n}}$ и бъде центриран в елемент с координати (u, v) , а елементите на този прозорец са подредени в матрицата $\mathbf{Z}_{u,v} \in \mathbb{R}^{N_{\bar{n}}^2 \times 3}$. Като първа стъпка се изчислява ковариационната матрица на данните от прозореца, като преди това те се центрират както следва (виж Фиг.4.5):

$$\begin{aligned} \mathbf{x}_{\mu} &= \frac{1}{N_{\bar{n}}^2} \mathbf{Z}_{u,v} \mathbf{e}, \\ \bar{\mathbf{Z}}_{u,v} &= \mathbf{Z}_{u,v} - \mathbf{x}_{\mu} \mathbf{e}^T, \\ \mathbf{C}_{uv} &= \frac{1}{N_{\bar{n}}^2} \bar{\mathbf{Z}}_{u,v}^T \bar{\mathbf{Z}}_{u,v}, \end{aligned} \tag{4.8}$$

където \mathbf{e} е единичният вектор. Така изчислената ковариационна матрица се разлага по собствени вектори и стойности:

$$\mathbf{C}_{uv} = \mathbf{S}^T \Lambda \mathbf{S}, \tag{4.9}$$



Фигура 4.6: Блок-схема на алгоритъм за класификация на 3D изображения на лица с множество ядра - фаза Тест.

където \mathbf{S} са собствените вектори, които са ортонормални, а Λ е диагонална матрица съдържаща собствените стойности подредени по намаляващ ред. За нормала $\vec{\mathbf{n}}_{u,v}$ към прозореца с координати (u, v) се взима собственият вектор, който съответства на най-малката собствена стойност:

$$\vec{\mathbf{n}}_{u,v} = \arg \min_{\mathbf{S}} \Lambda. \quad (4.10)$$

Така извлечени нормалните за всички елементи от изображението се подреждат в матрицата $\vec{\mathbf{N}}$ и се подават за по-нататъшна обработка.

4.3.3 Извличане на признаци чрез Locally Adaptive Regression Kernels

LARK е алгоритъм за извличане на признаци, предложен от [171]. С този алгоритъм се измерват разстоянията между точки (пиксели) чрез геодезичното разстояние, т.е. разстоянието определено от многообразието на сигнала. Това се постига, като се използва локалното сходство базирано на градиентите. Следователно, чрез този алгоритъм се улавя локалната геометрична структура на лицето, което е от значение за 3D разпознаването на лица. Математическите означения в тази точка са идентични с тези на Фиг. 4.5 и фигурата може да бъде използвана за референция.

Алгоритъмът работи върху прозорци от изображението, които са центрирани в елемент с координати (u, v) . За всеки един елемент от прозореца се измерва разстоянието между него и централния елемент, което разстояние е най-близкият път между тези два елемента лежащ на повърхнината, тоест геодезичното разстояние. Нека повърхнината на изображението е параметризирана по (u, v) , $S(u, v) = \{u, v, z(u, v)\}$. Тогава квадрата на елементарното разстояние между две точки е сума от елементарните разстояния на квадрат, по всяка координата: $ds^2 = du^2 + dv^2 + dz^2$. Диференцирайки $z(u, v)$, и изчислявайки ds^2 , се получава [171]:

$$ds^2 = \Delta \mathbf{u}^T \mathbf{C}^L \Delta \mathbf{u} + \Delta \mathbf{u}^T \Delta \mathbf{u}, \quad (4.11)$$

където $\Delta \mathbf{u} = [du, dv]^T$ е диференциала по отношение на независимите променливи. Матрицата \mathbf{C}^L се нарича структурен тензор и представлява ковариационната матрица на градиентите в локална област и се изчислява чрез:

$$\mathbf{C}^L = \begin{bmatrix} z_u^2 + 1 & z_u z_v \\ z_u z_v & z_v^2 + 1 \end{bmatrix}, \quad (4.12)$$

където z_u и z_v са производните по u и v съответно. Геодезичното разстояние не се използва директно, а се претегля през експоненциална функция, за да се намалят теглата по нелинеен закон с увеличаване на разстоянието. Алгоритъмът за извличане на признаците LARK [171], е представен на Алг. 1.

Algorithm 1: Алгоритъм за извличане на признаците LARK

Вход: \mathcal{D} - Растеризирано изображение
Вход: N_{LARK} - Размер на прозореца
Изход: \mathbf{K}^L - Признаци LARK за изображението

- 1 **while** Край на Изображение **do**
- 2 $\Omega_{u,v} \in \mathcal{D}$ - центриране на прозорец в (u, v)
- 3 Изчисляване на структурния тензор (4.12) за $\Omega_{u,v}$
- 4 $\bar{\mathbf{C}}^L = \sum_{(u,v) \in \bar{\Omega}_{u,v}} \begin{bmatrix} z_u(u, v)^2 & (z_u z_v)(u, v) \\ (z_u z_v)(u, v) & z_v(u, v)^2 \end{bmatrix}$ - усредняване
- 5 Регуляризиране на $\bar{\mathbf{C}}^L$ чрез претегляне на собствените стойности
- 6 $K_{u,v}^L = \exp(-\Delta \mathbf{u}^T \bar{\mathbf{C}}^L \mathbf{u})$ - претегляне с експонента
- 7 $\mathbf{K}^L \leftarrow K_{u,v}^L$ - попълване на матрица

Стъпка 4 е необходима, за да се намали шума, който се съдържа в частните производни. Усредняването става като се сумират структурните тензори поелементно, в околност $\bar{\Omega}_{u,v}$, която е по-малка от $\Omega_{u,v}$. Стъпка 5 от алгоритъма е описана подробно в [171], но основната идея се свежда до претегляне на собствените стойности на $\bar{\mathbf{C}}^L$, за да се избегне ранг 1 на матрицата. Резултатът от всеки прозорец се представя като вектор стълб и се попълва в матрицата \mathbf{K}^L . Тази матрица е с размерност $|\mathcal{D}| \times N_{LARK}^2$, където $|\mathcal{D}|$ е броят на елементите на изображението. След като изчислят признаците LARK се прилага PCA за всяко изображение независимо от останлите, за да се намали размерността, и резултатът \mathbb{K}^L се подава за по-нататъшна обработка.

4.3.4 Класификация на 3D лица чрез множество ядра

В методите за машинно обучение реализирани по метода с ядрото, най-важния компонент е матрицата на Грам, която се изчислява за всяка двойка от елементи от обучаващата извадка с предварително избрано ядро $k(\cdot, \cdot)$. Следователно, избора на правилно ядро, с подходящи параметри, е критичен за постигането на добри резултати. Един от подходите за избор на ядро и параметри, който стъпва на аналитични аргументи, е обучение чрез множество от ядра [173]. при него изходната матрица на Грам се изчислява като изпъкнала комбинация от базови ядра:

$$K^R(\cdot, \cdot) = \sum_{m=1}^{K_M} d_m K_m(\cdot, \cdot), \quad d_m \geq 0, \quad \sum_{m=1}^{K_M} d_m = 1, \quad (4.13)$$

където K_M е броят на базовите ядра, d_m са коефициентите на изпъкналата им комбинация, а $K_m(\cdot, \cdot)$ са базовите ядра. Използвайки резултатното ядро $K^R(\cdot, \cdot)$, според теоремата на представителя (виж т. 1.3.1), всяка функция в хилбертовото пространство с възпроизвеждащо ядро може да се представи с:

$$f^*(\cdot) = \sum_{i=1}^M \alpha_i K^R(\cdot, x_i) + \beta, \quad (4.14)$$

където $\alpha_i \in \mathbb{R}$, а $x_i, i = 1, \dots, M$ е обучаващата извадка. Следователно, когато се използва множество ядра, обучението се свежда до намиране на коефициентите α_i и d_m . Така дефинирана задачата е изпъкнала, но въпреки това тя е трудна за решаване, защото не е гладка, тоест първата производна е прекъсната [174].

В текущата работа за разпознаване на 3D изображения на лица се използва формулировка на SVM реализирана с множество ядра, наименувана SimpleMKL и предложена в [174]. Предложените множество ядра се състоят от две групи, всяка предназначена за различните признаци описани в предходните две точки. Всяка група се състои от един тип ядро, но реализирано с различни параметри. Първата група използва множество ядра, които са подходящи за нормалите към повърхнината, а втората подходящи за LARK признаците. От Фиг. 4.5 се вижда, че нормалите са означени с \vec{N} , което представлява матрица с размери $|\mathcal{D}| \times 3$. Вземайки предвид, че нормалите към повърхнина са с дължина единица е подходящо да се използва полиномно ядро, защото при него се извършва скаларно произведение между признаците в явен вид, тоест сходството между две нормали представлява косинуса на ъгълът между тях. Нека \vec{N}_i и \vec{N}_j са нормалните за i -тото и j -тото изображения съответно. Тогава полиномното ядро се изчислява по следният начин (виж също Фиг. 4.5):

$$k^p(\vec{N}_i, \vec{N}_j) = (\eta \text{tr}\{\vec{N}_i \vec{N}_j^T\} + \zeta)^3, \quad (4.15)$$

където η и ζ са параметрите на полиномното ядро, и както се вижда, то е избрано да бъде от ред 3. За да се постигне оптимален избор на параметрите на полиномното ядро, то се дефинира като множество ядра с параметри $\{\eta_n, \zeta_n\}_{n=1}^{Pp}$.

Признаците LARK нямат толкова явна интерпретация като нормалните, поради това, най-добрият избор е гаусовото ядро заради неговата универсалност [175]:

$$k^g = \exp(-\kappa \|K_i^L - K_j^L\|), \quad (4.16)$$

където κ е параметърът на гаусовото ядро. Аналогично на полиномното, се избират няколко различни стойности на параметъра, за да се генерира множество ядра $\{\kappa\}_{n=1}^{Pg}$. Така се задават двете групи от множество ядра, с които се дефинира резултантното ядро, което се използва в алгоритъма SVM:

$$k^R(\cdot, \cdot) = \sum_{m=1}^{Pg} d_m k^g + \sum_{n=1}^{Pp} d_n k^p, \quad (4.17)$$

където коефициентите d_m съответстват на гаусовите ядра, а d_n съответстват на полиномиалните ядра. Използвайки резултантното ядро се построява матрицата на Грам K^R .

Използваният алгоритъм, предложен в [172], замества регуляризацията със смесената норма с L_2 нормата, а разредеността на множеството ядра се контролира с L_1 нормата. Тази формулировка задава изпъкнала и гладка задача, което позволява използването на стандартни градиентни методи, при което формулировката на SVM алгоритъма с множество ядра се свежда до стандартната формулировка

с едно ядро, което е резултатното:

$$\max_{\alpha, d_m, d_n} -\frac{1}{2} \sum \alpha_i \gamma_i K_{i,j}^R \gamma_j \alpha_j + \sum \alpha_i \quad (4.18)$$

При обучението се намират оптималните опорни вектори и теглата на множеството ядра α, d_m, d_n , които се използват във фазата тест. Решаването на (4.18) се извършва на итерации, при които се решава стандартният SVM оптимизационен проблем, последван от решаването на разредената задача за теглата на множеството ядра, което може да се извърши със симплекс метода.

След като са намерени опорните вектори и оптималните параметри на множеството ядра, класификацията се извършва сходно на стандартния SVM алгоритъм, но използвайки резултатното ядро:

$$\hat{\gamma} = \text{sgn}(\sum \alpha_i \gamma_i K^R + b). \quad (4.19)$$

4.4 Основни резултати и изводи

Основните резултати от работата в тази глава са следните:

1. Разработен е алгоритъм за разпознаване на 3D изображения на човешки лица, базиран на дълбочинно изображение, състоящ се от:
 - Предварителна обработка, адресираща грешките възникващи при процеса на сканиране, като шум и липсващи данни.
 - Извличане на региона на лицето.
 - Преобразуване на облака от точки на лицето в дълбочинно изображение, което е с нормализиран динамичен диапазон.
 - Извличане на локални признаци, представени от локалните гаусова и средна кривини, които са инварианти към ротация и трансляция.
 - Комбинирано намаляване на признаците с PCA и LDA, предоставящо дискриминативност на признаците.
 - Класификация с алгоритъм SVM, реализиран с ядро.
2. Разработен е алгоритъм за разпознаване на 3D изображения на лица с множество ядра, включващ:
 - Предварителна обработка адресираща грешките възникващи при процеса на сканиране, като шум и липсващи данни, и растеризация по хоризонтална и вертикална координати.
 - Извличане на признаци, представени чрез нормалите към повърхнината на лицето, изчислени за околност на всеки елемент от нея, които признаци кодират скоростта на промяна на повърхнината и са инвариантни към ротация и трансляция.
 - Извличане на признаци представени чрез LARK, които кодират геодезичното разстояние между елементите на изображението, в околност на избран елемент. Признаците са инвариантни към трансляция, и частично

инвариантни към ротация, и се явяват допълващи към нормалите, защото нормалите използват диференциална компонента, докато LARK използват абсолютни стойности, като разстояние.

- Построяване на множество ядра в две групи, първата група предназначена за нормалите, а втората за признаците LARK. Това множество включва и различни параметри на ядрата.

- Използване на множеството ядра в класификатора SVM за осъществяване на класификацията и избор на оптимални параметри на ядрата.

Изводите от работата по тази глава са следните:

1. Локалните признаци базирани на кривината на повърхнината са информативни за задачата за разпознаване на 3D изображения на лица.
2. Използването комбинация на нормалите и признаците LARK е ефективно за задачата за разпознаване на лица;
3. Множеството ядра адресират проблемите, свързани с избора на параметри на ядрата, увеличаване на капацитета на обучение на класификатора, и комбинацията на разнообразни признаци.

4.5 Приноси към Глава 4

Научните приноси са:

1. Разработен е метод, базиран на допълващи се геометрични признаци и комбинацията им с множество ядра, използвани като резултантно ядро в класификатора SVM. Този принос е към Глава 4, точка 4.3.

Научно-приложните приноси са:

1. Разработен е подобряващ дискриминативността алгоритъм за разпознаване на 3D лица, използващ геометрични признаци, базирани на кривината на формата и проекцията им в подпространство. Този принос е към Глава 4, точка 4.2.
2. Разработен е алгоритъм за разпознаване на 3D лица, базиран на допълващи се геометрични признаци и оптималната им комбинация чрез множество ядра. Допълващите се геометрични признаци увеличават информативността, а оптималната им комбинация увеличава дискриминативността на класификатора. Този принос е към Глава 4, точка 4.3.

Приложните приноси са:

1. Програмна реализация и експериментални изследвания на алгоритъм за разпознаване на 3D лица, използващ извличане на геометрични признаци и проекция в подпространство. Постигнат коефициент на разпознаване от 96.64%, използвайки стандартна база данни от лица. Този принос е към Глави 4 и 5, точки 4.2 и 5.3.1.

2. Програмна реализация и експериментални изследвания на алгоритъм за разпознаване на 3D лица, използващ допълващи се геометрични признаци и интегрирано в SVM множество ядра. Постигнат коефициент на разпознаване от 94.90%, подобряващ докладваните в литературата, използвайки стандартна база данни от лица. Този принос е към Глави 4 и 5. Този принос е към Глави 4 и 5, точки 4.3 и 5.3.2.

Алгоритмите за разпознаване на 3D лица, разработени в тази глава, са докладвани в следните научни публикации:

- [D4] **K. Tonchev**, A. Manolova, I. Paliy, "Comparative analysis of 3d face recognition algorithms using range image and curvature-based representations." In 2013 *IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS)* (Vol. 1, pp. 394-398). IEEE. DOI: 10.1109/IDAACS.2013.6662714
- [D5] **K. Tonchev**, "3D Face Recognition with Multiple Kernel Learning." *WSEAS International Conference. Proceedings. Recent Advances in Computer Engineering Series*. No. 7. WSEAS, 2012. ISBN: 978-1-61804-134-0

5 Експериментални резултати от разработените алгоритми

В настоящата глава са описани резултатите, получени от тестването на създадените алгоритми, които са описани в предходните глави на дисертацията за фазата Тест. Изследванията относно ефективността на разработените алгоритми са проведени с популярни бази данни, съдържащи изображения на лица, заснети в контролирани условия или събрани на случаен принцип от Интернет. Експериментите са извършени на работна станция HP Z600 със следните параметри:

- Процесор Intel Xeon Quad-Core 2.80GHz;
- 6GB DDR2 1400MHz оперативна памет;
- Твърд диск с обем 1.5TB;

Използваната операционна система е Linux, дистрибуцията Ubuntu.

5.1 Експериментални изследвания на алгоритмите към Глава 2

5.1.1 Експериментални резултати от теста на алгоритъма за определяне на възраст в групи

В тази точка са представени експериментални резултати от тестовете за ефективност на алгоритъма за определяне на възраст в групи, предложен в т. 2.2. Реализацията на отделните алгоритми, които се използват в предложеният алгоритъм, е както следва:

- Алгоритъмът за регистриране на лица, PCA и SR са програмирани на MATLAB [176].
- Алгоритъмът за детектиране на лица е реализиран на C++ в библиотеката OpenCV [177].
- Класификаторът на лица с CNN е реализиран на C++.
- Алгоритъмът SVM е реализиран на C++ в библиотеката libsvm [178].

Алгоритъмът SVM е реализиран в библиотеката libsvm и използва метода Sequential Minimal Optimization [179] за решаване на оптимизационната задача. В библиотеката също така са включени нормализация на данните, както и крос-валидационна процедура за избор на регуляризационния параметър и параметъра на ядрото.



ФИГУРА 5.1: Примерни изображения от базата данни FG-NET.

Нормализацията и избора на оптималните параметрите, предоставени в тази библиотека, се използват в текущата работа. Разлагането по собствени стойности и вектори, което се ползва в алгоритмите PCA и SR, се извършва от функцията *eig* предоставена в пакета MATLAB.

За да се тества предложения алгоритъм за определяне на възраст в групи се използва базата данни FG-NET [95]. Тя се състои от 1002 цветни и полутонови изображения, на лица на 82 индивида, със средно по 12 изображения на индивид, всяко от които е с аотирана възраст. Възрастовите граници на индивидите в библиотеката са от 0 до 69 години. Към всяко изображение са предоставени и 82 ръчно аотирани точки от лицето. Изображенията са заснети в неконтролирана среда с вариращо качество и големи вариации на позата. Голяма част от изображенията са сканирани от снимки. Това увеличава значително вариациите на данните и усложнява задачата за определяне на възраст. Примерни изображения от базата данни са дадени на Фиг. 5.1.

Предложеният в т. 2.2 алгоритъм е тестван за определяне на възраст в две групи (класа), деца и възрастни. В първата попадат изображенията на индивиди, които са до 16 годишна възраст, а във втората са изображенията аотирани с възраст от 17 до 69 години. Това разделяне е избрано, защото в първата група попадат малолетните, а във втората непълнолетните и пълнолетните. За изследва ефективността на алгоритъма е извършен сравнителен анализ между следните четири конфигурации:

1. Намаляване на размерността с PCA и класификация с SVM (PCA+SVM).
2. Намаляване на размерността с PCA и класификация с KSVM (PCA+KSVM).
3. Намаляване на размерността с PCA+SR и класификация с SVM (PCA+SR+SVM).

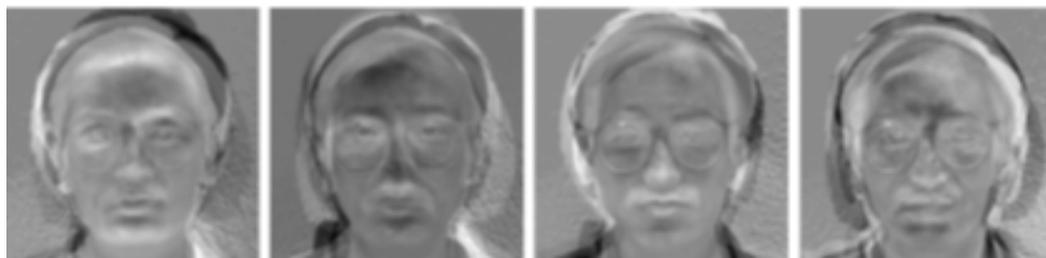
4. Намаляване на размерността с PCA+KSR и класификация с SVM (PCA+KSR+SVM).
5. Намаляване на размерността с PCA+SR и класификация с KSVM (PCA+SR+KSVM).

Първата конфигурация е базова и не използва метода с ядрото. Тя се използва за да се оцени, колко се увеличава ефективността с предложения алгоритъм, когато се използва метода с ядрото. Втората конфигурация използва намаляване на размерността с PCA, тоест линейно преобразуване, което отстранява основно шум, и нелинеен SVM класификатор с ядро (KSVM). Третата конфигурация използва вграждане на възрастово многообразие в подпространство чрез линейно преобразуване. В тази конфигурация се използва линеен SVM, защото се предполага, че подпространството в което SR проектира данните е запазило дискриминативността им. Аналогични са разсъжденията и за четвъртата конфигурация, но при нея се използва SR реализиран с ядро, поради което се очаква този алгоритъм да има най-голяма ефективност. Класификаторът на последната конфигурация работи с най-малък размер на данните, при което се очаква тази конфигурация да е компромисна по отношение на ефективност-бързодействие. Прилагането на метода за крос-валидация при определянето на точността на алгоритъма е некоректно, защото изображения на един и същи субект могат да са, както в обучаващата извадка, така и в тестовата. Това ще доведе до изместване на оценката при тест, защото алгоритъма е обучен с данни на субекта, който ще се тества. Поради това се използва процедурата *LOPO*, при която всички изображения от един субект се използват за тест, а останалите се използват за обучение.

Параметрите на алгоритъма са представени в Таблица 5.1. Параметрите, които са избрани на база на експерименти и с които се постигат най-добри резултати, са записан в таблицата като „експериментално”. Визуална представа за базисните функции комбиниращи преобразуванията PCA и SR може да се придобие от Фиг. 5.2, на която са показани водещите четири собствени вектора, изчислени при прилагане на алгоритъма за класификация във възрастови групи. Интересен извод,

ТАБЛИЦА 5.1: Параметри на алгоритъма за определяне на възраст в групи.

Параметър	Стойност
Размер на изображенията след детектора на лица и регистрация	40 × 50 пиксела
Праг на PCA за избор на главни компоненти, \mathcal{T}	0.95
Размерност на признаците след PCA	116 елемента
Размерност на признаците след SR	12 елемента (експериментално)
Размерност на признаците след KSR	2 елемента
Параметър на гаусовото ядро на SR и KSR, σ	7 (експериментално)
Регуляризиращ параметър на SR и KSR, β	0.05 (експериментално)
Регуляризиращ параметър на SVM и KSVM, β	0.8 (експериментално)
Параметър на ядрото на KSVM, t	12 (експериментално)



ФИГУРА 5.2: Примерен базис на PCA и SR базис, генериран с изображения на лица.

който може да се направи от фигурата е, че базисните функции могат да се интерпретират, като холистични високочестотни филтри, концентрирани около местата с най-много вариации като очите, контурите на лицето и устата.

Определянето на възраст е в две групи, което означава, че задачата е за класификация в два класа. Поради това, че се използва процедурата LOPO, не се използва Receiver Operating Characteristics (ROC) кривата [180] за определяне на ефективността на предложеният алгоритъм. Определянето на ефективността се базира на средният коефициент на разпознаване (Mean) и стандартното му отклонение (STD). Те се изчисляват, като се приеме, че коефициентът на разпознаване за всеки индивид от процедурата LOPO е случайна величина. При използване на LOPO процедурата с базата FG-NET, средно се падат по 900 изображения за обучение и 12 изображения за тест. Важно е да се отбележи, че предложият алгоритъм не е сравнен с резултатите на други, публикувани в литературата, защото към момента на публикуване на статията, няма публикации съдържащи сходни експерименти. Резултатите от експериментите са представени в Таблица 5.2.

ТАБЛИЦА 5.2: Коефициент на разпознаване (ефективност) на предложения алгоритъм за различните конфигурации.

Конфигурация	Mean[%]	STD[%]
PCA+SVM	74.8	± 16.39
PCA+KSVM	77.73	± 15.18
PCA+SR+SVM	76.90	± 15.78
PCA+KSR+SVM	77.67	± 14.0
PCA+SR+KSVM	77.8	± 15.96

Анализ на резултатите.

Както се очаква, базовата конфигурация има най-ниска ефективност, защото използва линейни методи, без да взима предвид многообразието на данните. Третата конфигурация също използва само линейни методи, но постига по високи резултати, защото алгоритъмът SR успява да запази дискриминативността на данните при вграждането на многообразието в подпространство. Тази конфигурация е за предпочитане, когато се цели бързодействие. Втората, четвъртата и петата конфигурации имат сходни резултати. Петата има най-висока ефективност, но вариацията на грешката е най-голяма, което означава, че клъстерите на класовете не са компактни. Четвъртата има най-ниска ефективност от изброените три, но има значително по-ниска вариация на разпознаването, поради което тя се счита

за най-ефективна. Ниската вариация на тази конфигурация се дължи на използването на метода с ядрото, за намаляването на размерността на данните с SR алгоритъма, защото с него се постига по-добро моделиране на многообразието на данните.

5.2 Експериментални изследвания на алгоритмите към Глава 3

5.2.1 Експериментални резултати от алгоритъма за разпознаване на изражения с избор на признаци в log-Gabor признаково пространство

В тази точка са представени експериментални резултати от тестовете за ефективност на алгоритъма за разпознаване на изражения, с избор на log-Gabor признаци, предложен в т. 3.2. Реализацията на отделните алгоритми, които се използват в предложеният алгоритъм е както следва:

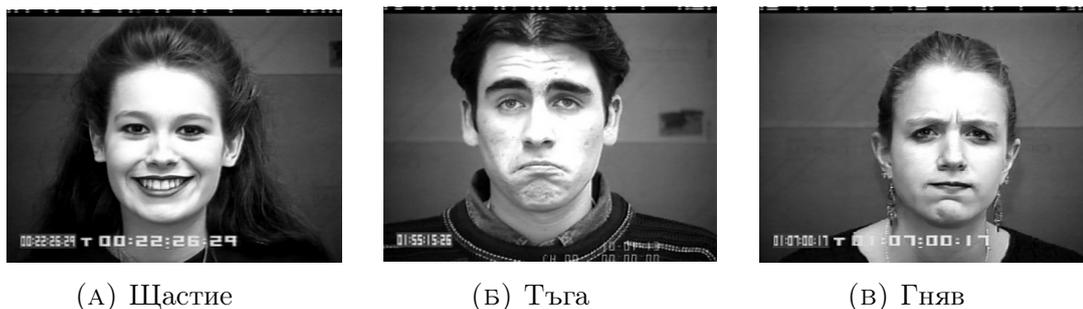
- Алгоритъмът за регистриране на лица е програмиран на MATLAB.
- Алгоритъмът за детектиране на лица е реализиран на C++ в библиотеката OpenCV [177].
- Класификатора на лица с CNN е реализиран на C++.
- Алгоритъмът за извличане на log-Gabor признаци е програмиран на C++, използвайки функционалности от библиотеката OpenCV.
- Алгоритъмът за избор на признаци е програмиран на MATLAB.
- Алгоритъмът SVM е реализиран на C++ в библиотеката libsvm.

Нормализацията на данните и избора на оптимални параметри, които се извършват в libsvm, се използват в текущата работа.

За определяне на ефективността на алгоритъма се използва базата данни Cohn-Kanade Extended Facial Expression Database (CK+) [122]. Тази база данни е предназначена за разработване и изследване на алгоритми за разпознаване на лицеве изражения. Тя съдържа 2D полутонови изображения на лица на 123 души, като общо са налични 593 видео последователности. В базата данни има налични изображения на лица на 118 индивида, които са с анотирани изражения и се използват за обучение и тест на предложеният алгоритъм. Израженията налични в базата данни са 7 класа (типа): „гняв”, „презрение”, „отвращение”, „страх”, „щастие”, „тъга” и „изненада”. Разпределение на броя на изображенията за всеки клас, използвани за експериментите, е даден в Таблица 5.3.

ТАБЛИЦА 5.3: Брой изображения за различните изражения в CK+.

Гняв	Презрение	Отвращение	Страх	Щастие	Тъга	Изненада
45	18	59	25	69	28	83



ФИГУРА 5.3: Примерни изображения от базата данни СК+.

Тук е важно да се отбележи, че броя на изображенията в различните класове не е еднакъв и дори има класове, като „презрение”, в които изображенията са близо три пъти по-малко от класа „щастие”. Този дисбаланс води до изместване на оценката при разпознаването, и в общия случай, се прибегва до подходи за балансиране на броя на изображенията по класове. Такъв подход няма да бъде използван за оценка на ефективността на предложеният алгоритъм поради две причини: увеличаване на изображенията чрез добавяне на изкуствено генерирани вариации не е релевантно, защото алгоритъма за регистрация ще компенсира тези вариации; броят на изображенията в извадката е малък, поради което, допълнително намаляване не би било от полза. Друг аргумент за запазване на изображенията по класове е, че те са генерирани в контролирани условия от актьори, и съответно са с високо качество за задачата за разпознаване на изражения.

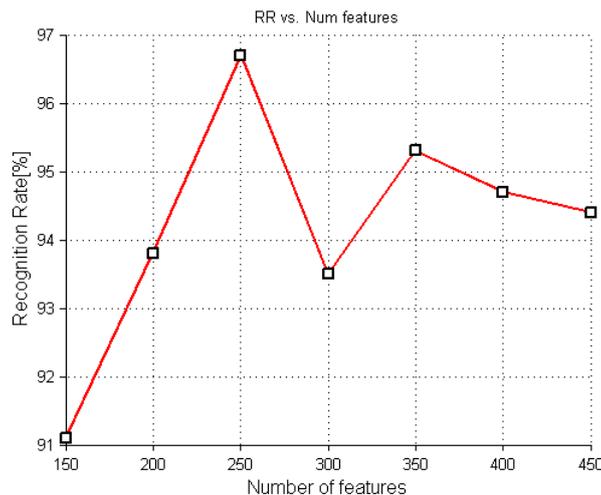
Примерни изображения от базата данни са представени на Фиг. 5.3.

Предложеният в т. 3.2 алгоритъм за разпознаване на изражения е тестван в 7 класа. Тестът се извършва по процедурата LOPO, като изображенията на израженията на един индивид се използват за тест, а останалите за обучение. За всяка итерация от процедурата LOPO се използват 861 изображения за обучение и 7 за тест, колкото са израженията.

Параметрите на алгоритъма са представени в Таблица 5.4. Параметрите които са избрани на база на експерименти, и с които се постигат най-добри резултати, са

ТАБЛИЦА 5.4: Параметри на алгоритъма за разпознаване на изражения с избиране на log-Gabor признаци.

Параметър	Стойност
Размер на изображенията след детектора на лица и регистрация	60 × 50 пиксела
Брой на скалите на log-Gabor	3
Брой на ориентациите на log-Gabor	5
Брой на избраните признаци	250 елемента (експериментално)
Параметър на гаусовото ядро на SR, σ	0.1 (експериментално)
Регуляризиращ параметър на SR, β	2 (експериментално)
Регуляризиращ параметър на KSVM, β	10 (експериментално)
Параметър на ядрото на KSVM, t	0.5 (експериментално)



ФИГУРА 5.4: Разпознаването на израженията като функция от броя на признаците избрани от алгоритъма за избор на признаци.

записани в таблицата като „експериментално”. Размерността на признаковия вектор, който се подава на алгоритъма за избора на признаци, тоест след \log -Gabor, е $3 \times 5 \times 60 \times 50 = 45000$. Това означава, че след избора на признаци, размерността е намалена 180 пъти до 250 елемента. Изборът на признаците се извършва като се изпълни алгоритъмът за различни стойности на броя на признаците в интервала $[150 - 450]$, със стъпка 50, след което се избере броят на признаците, за които стойността на разпознаване е най-голяма. Резултатите от този експеримент са представени на Фиг. 5.4. При изпълнението на този експеримент обаче, се използва стандартна процедура крос-валидация със стъпка 10. Както се вижда от фигурата, броят на признаците, за които коефициента на разпознаване е най-голям е 250 и именно този брой е използван за определяне на ефективността на алгоритъма.

Поради това, че се използва процедурата LOPO, се изчисляват средната стойност (Mean) и стандартното отклонение (STD) на коефициента на разпознаване, разгледан като случайна величина. Средната стойност на разпознаване на алгоритъма е сравнена с алгоритмите предложени в [121], [120] и [181]. Тези алгоритми са избрани за сравнение, защото използват същата база данни и същата процедура за определяне на ефективността. Средната стойност на разпознаване и стандартното отклонение при проведените тестове, както и разпознаването на избраните за сравнение алгоритмите, са представени в Таблица 5.5.

ТАБЛИЦА 5.5: Резултати от експериментите за определяне на ефективност на предложения алгоритъм и сравнението му с други.

Алгоритъм	Разпознаване[%]
Ou et al. [121]	80
Bashar et al. [181]	92.2
Shan et al. [120]	92.6
Предложеният алгоритъм	96.7

Анализ на резултатите.

Резултатите от сравнителният анализ показват, че предложеният алгоритъм е по-ефективен от други докладвани в литературата, при сходни или еднакви условия на тест. На база на тези резултати може да се направи извод, че избора на признаци log-Gabor е ефективен, и избраните признаци са информативни. Също така, влиянието на „проклятието на размерността” е намалено, което се доказва чрез постигнатото намаляване на размерността от 180 пъти.

5.2.2 Експериментални резултати от алгоритъма за разпознаване на изразения на лица за големи обеми от данни, реализиран по метода с ядрото

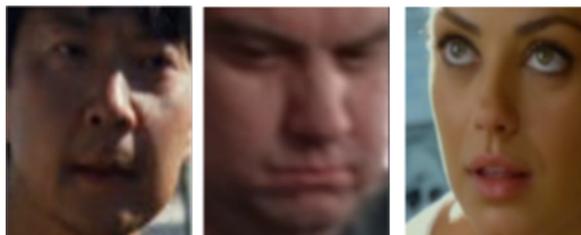
В тази точка са представени експериментални резултати от тестовете за ефективност на алгоритъма за разпознаване на изразения с големи обеми от данни, предложен в т. 3.3. Реализацията на отделните алгоритми, които се използват в предложеният алгоритъм е както следва:

- Алгоритъмът за регистриране на лица е програмиран на MATLAB.
- Алгоритъмът за детектиране на лица е реализиран на C++ в библиотеката OpenCV [177].
- Класификатора на лица с CNN е реализиран на C++.
- Алгоритъмът за детектиране на ключовите точки на лицето SDM, е реализиран на MATLAB.
- Обучението на алгоритъма за класификация с мултиномиална регресия и множество ядра е реализиран на MATLAB.
- Алгоритъмът за оптимизация, Stochastic Gradient Descent, се използва като готова функция предоставена в пакета на MATLAB.

Алгоритъмът SDM се обучава преди използването му в предложения алгоритъм. Коефициентите на обучения SDM алгоритъм са предоставени от авторите му. Те са го обучили с 884 изображения на лица, анотирани с ключовите точки, и са го тествали с 245 изображения на лица.

За оценяване на ефективността на алгоритъма се използва базата данни Static Facial Expressions in the Wild (SFEW), която е част от предизвикателството EmotiW 2015 [128]. Тази база данни съдържа 1590 изображения на човешки лица, включващи произволен фон, избрани от видео последователности. Избора е извършен чрез алгоритъм за извличане на ключови кадри от видео. Базата е допълнително разделена на 880 изображения за обучаваща извадка, 383 за валидация и 372 изображения за тест. Примерни изображения от база данни са дадени на Фиг. 5.5. От фигурата се вижда, че изображенията са с ниско качество и са заснети в неконтролирани условия, което прави задачата за разпознаване на изразения предизвикателна.

Предложеният в т. 3.3 алгоритъм за разпознаване на изразения е тестван в 7 класа, представляващи следните 7 лицеви изразения, предоставени в базата



ФИГУРА 5.5: Примерни изображения от базата данни SFEW 2.0.

данни: „гняв”, „отвращение”, „страх”, „щастие”, „неутрално”, „тъга” и „изненада”. Ефективността на алгоритъма се оценява като се използва средният коефициент на разпознаване за тестовата и валидационна извадки. Тези оценки са избрани за да съвпадат с протокола на предизвикателството EmotiW, което ще позволи алгоритъма да бъде сравнен с други докладвани в литературата. Алгоритъмът се обучава чрез разделяне на случайните признаци на 40 блока, всеки от тях състоящ се от 25000 признака. Обучението се провежда паралелно върху всеки блок чрез **SAG**, но само върху обучаващата извадка.

Използваните в експерименталната постановка ядра са гаусовото и лапласовото. Параметрите и на двете ядра са избрани експериментално да бъдат следното множество $\{0.1, 0.5, 1, 2, 4\}$. Това води до множество от 10 ядра – 5 гаусови и 5 лапласови. Този избор на параметрите на ядрата избягва процедура за определяне на оптимални параметри чрез обхождане на пространството им, която процедура е много времеемка. Броят на случайните признаци е избран експериментално да бъде 1,000,000. Параметрите на алгоритъма са обобщени в Таблица 5.6.

Средната стойност на разпознаване на алгоритъма, за тестовата и валидационни извадки, е сравнена с алгоритмите предложени в [129] и [130]. Тези алгоритми са избрани за сравнение, защото използват същата база данни и същата процедура за определяне на ефективността. Първият от тях е базов метод за сравнение,

ТАБЛИЦА 5.6: Параметри на алгоритъма за разпознаване на изражения, използвайки апроксимация на метода с ядрото за големи обеми от данни.

Параметър	Стойност
Размер на изображенията след детектора на лица и регистрация	100 × 100 пиксела
Размер на прозореца за извличане на SIFT	32 × 32 (експериментално)
Итерации на сходимост на SDM	5
Параметри на гаусовото множество ядра	$\{0.1, 0.5, 1, 2, 4\}$ (експер.)
Параметри на лапласовото множество ядра	$\{0.1, 0.5, 1, 2, 4\}$ (експер.)
Брой ядрата в множеството ядра	10
Брой признаци апроксимиращи ядрото	1 000 000 (експер.)
Брой блокове за паралелно обучение	40 (експер.)
Брой признаци за блок за паралелно обучение	25000 (експер.)

който също се използва като базов в предизвикателството EmotiW и е реализиран със съвременна конволюционна невронна мрежа. Вторият също използва CNN, но преди това извлича локалните признаци LBP. Изчислената средна стойност на коефициента на разпознаване на алгоритъма, както и тези на алгоритмите, докладвани в литературата, са представени в Таблица 5.7.

ТАБЛИЦА 5.7: Резултати от експериментите за определяне на ефективност на предложения алгоритъм и сравнението му с други.

Алгоритъм	Валидация[%]	Тест[%]
Алгоритъмът предложен от [129]	68.4	47.6
Алгоритъмът предложен от [130]	51.74	54.56
Предложеният алгоритъм	64.4	53.6

Анализ на резултатите.

Важно е да се отбележи, че предложеният метод дава сравними резултати със съвременните конволюционни невронни мрежи и дори може да се разглежда като „плитка” (shallow) мрежа от ядра. Коефициентът му на разпознаване е по-висок от базовия метод на предизвикателството EmotiW, и сходен, но по-нисък от този на алгоритъма, предложен в [130]. Експерименталните резултати показват, че използваните множество ядра, апроксимирани със случайни признаци, могат да бъдат сравнени по степен на разпознаване със съвременните методи за разпознаване на изражения на лица, като в същото време поддържат ниска изчислителна сложност. Разделянето на случайните признаци на блокове позволява паралелно обучение и използване на модерни паралелни изчислителни платформи.

5.3 Експериментални изследвания на алгоритмите към Глава 4

5.3.1 Експериментални резултати от алгоритъма за разпознаване на 3D лица чрез представяне с дълбочинно изображение

В тази точка са представени експериментални резултати от тестовете за ефективност на алгоритъма за разпознаване на 3D лица, предложен в т. 4.2. Реализацията на отделните алгоритми, които се използват в предложения алгоритъм е както следва:

- Алгоритмите за отстраняване на точков шум и запълване на липсващи елементи са реализирани на MATLAB.
- Алгоритъмът за извличане на лицевият регион и регистрацията му са реализирани на MATLAB.
- Алгоритъмът за изчисляване на средната и гаусова кривини са реализирани на MATLAB.
- Алгоритмите PCA и LDA са реализирани на MATLAB.



ФИГУРА 5.6: Примерна извадка от 3D сканирани лица от базата данни SHREC.

- Класификаторите които се използват за сравнителен анализ са реализирани в Recognition Toolbox PRTools, MATLAB.
- Алгоритъмът SVM е реализиран на C++ в библиотеката libsvm.

Нормализацията на данните и избора на оптимални параметри, които се извършват в libsvm, се използват в текущата работа.

Предложения алгоритъм за разпознаване на 3D лица е тестван с базата данни SHape REtrieval Contest 2008 (SHREC) [158]. Всички субекти в базата данни са сканирани с лазерен скенер за разстояние Minolta Vi-700. За всеки индивид са извършени седем сканирания, две - с неутралното изражение, докато останалите са с различни изражения. Така се получават общо 61 сканирания, със 7 изображения за всеки индивид (общо 427 изображения), а именно – две „фронтални“, едно „поглед нагоре“, едно „поглед надолу“, едно „усмивка“, едно „смях“, и едно „произволно изражение“. Примерни изображения от базата данни са представени на Фиг. 5.6.

За да се изследва предложени в т. 3.2 алгоритъм, е направен сравнителен анализ на алгоритъма реализиран с различни класификатори. Класификаторите които са използвани са следните:

- Класификатор по най-малки квадрати (Fisher classifier).
- Много-класов квадратичен класификатор (QDC).
- Наивен бейсов класификатор.
- Класификатор по най-близките съседи (k-NN).
- SVM с линейно ядро (SVM k^{lin}).
- SVM с гаусово ядро (SVM k^{rbf}).

Два експеримента за оценка на ефективност са извършени, използвайки различни изражения на лицето. В първият експеримент, *Експеримент 1*, за обучение са използвани по две 3D изображения от индивид с неутрално изражение на лицето, а останалите за тест. За вторият експеримент, *Експеримент 2*, за обучение също са използвани по две 3D изображения от индивид, едното с изражение „усмивка“, а другото е избрано на случаен принцип. Останалите изображения са използвани за тест. Така като резултат, за *Експеримент 1*, алгоритъма се обучава със 122 изображения и се тества с 305. За *Експеримент 2*, алгоритъма се обучава със 183 изображения и се тества с 244. С тези два експеримента се цели да се установи

устойчивостта на работа на алгоритъма при наличие на изражения на лицето, когато се използват за обучение.

Параметрите на алгоритъма с които са извършени тестовете са представени в Таблица 5.8. Размерността на признаковият вектор, който се подава на PCA е $64 \times 64 = 4096$ елемента. Това означава, че PCA намалява размерността приблизително 36.5 пъти.

ТАБЛИЦА 5.8: Параметри на алгоритъма за разпознаване на 3D лица с дълбочинно изображение.

Параметър	Стойност
Размер на радиуса на сферата за извличане на лице	75мм
Размер на медианния филтър за точков шум	5×5
Размер на дълбочинното изображение след растеризация	64×64 пиксела
Праг на PCA за избор на главни компоненти, \mathcal{T}	0.95
Размерност на признаците след PCA	112 елемента

За определяне на точността се използва средният коефициент на разпознаване, който представлява средната стойност на разпознаването за всички Сравнителният анализ на предложения алгоритъм и реализацията му, с гореспоменатите класификатори, за двата проведени експеримента, е представена в Таблица 5.9.

ТАБЛИЦА 5.9: Резултати от експериментите на предложения алгоритъм, реализиран с различни класификатори и различни обучаващи извадки.

Класификатор	Разпознаване[%], Експеримент 1	Разпознаване[%], Експеримент 2
Fisher classifier	63.93	69.22
QDC	76.38	83.65
Наивен бейсов	72.68	79.22
k-NN	78.78	85.16
SVM k^{lin}	69.56	74.87
SVM k^{rbf}	87.12	91.56

Анализ на резултатите.

Сравнителният анализ на предложения алгоритъм, с използването на различни класификатори показва, че използването на класификатор, реализиран с ядро дава най-добри резултати. Друг извод, който може да се направи е, че ефективността пада когато алгоритъмът е обучен с неутрални изражения на лицето, а се тества с лица с изражения. Това се вижда от по-ниската ефективност, която се получава в *Експеримент 1*.

5.3.2 Експериментални резултати от алгоритъма за разпознаване на 3D лица с комбиниране на допълващи се локални признаци чрез множество ядра

В тази точка са представени експерименталните резултати от тестовете за ефективност на алгоритъма за разпознаване на 3D лица, реализиран с множество ядра и допълващи се признаци, предложен в т. 4.3. Реализацията на отделните алгоритми, които се използват в предложения алгоритъм е както следва:

- Алгоритмите за отстраняване на точков шум и запълване на липсващи елементи са реализирани на MATLAB.
- Алгоритъмът за извличане на лицевият регион и регистрацията му са реализирани на MATLAB.
- Алгоритъмът за изчисляване на нормалите към повърхнината е реализиран на MATLAB.
- Алгоритъмът за изчисляване на признаците LARK е реализиран на MATLAB.
- Алгоритъмът реализиращ множеството ядра е реализиран на MATLAB.

Предложения алгоритъм за 3D разпознаване на лица е тестван с базата данни SHape REtrieval Contest 2008 (SHREC) [158]. Всички субекти в базата данни са сканирани с лазерен скенер за разстояние Minolta Vi-700. На всеки субект са направени 7 сканирания, две от които са с неутрално изражение на лицето, а останалите с различни изражения на лицето. Сканирани са 61 субекта и общо 427 лицеви сканирания. Примерни изображения от базата данни са представени на Фиг. 5.7.



ФИГУРА 5.7: Примерна извадка от 3D сканирани лица от базата данни SHREC.

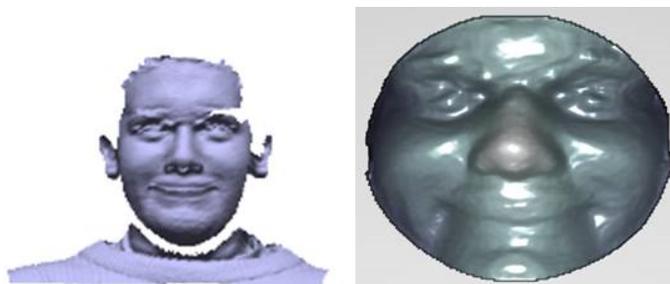
За да се изследва предложеният в т. 4.3 алгоритъм, е направен сравнителен анализ на ефективността на алгоритъма. Два експеримента са извършени използвайки различни обучаващи извадки. За първият експеримент, *Експеримент 1*, са използвани по две 3D изображения от индивид за обучение, а за вторият, *Експеримент 2*, по 3. Тези изображения са избрани на случаен принцип, което означава, че може да включват и неутрални, и лица с изражения.

Параметрите на алгоритъма, с които са извършени тестовете, са представени в Таблица 5.10. Параметрите, които са избрани експериментално, са маркирани в таблицата с експериментално.

На Фиг. 5.8 е представено изображение на извлеченото и нормализирано лице, преди етапа на извличане на признаци. От фигурата може да се забележи, че лицето е без области от липсващи елементи, както и изчистен точков шум.

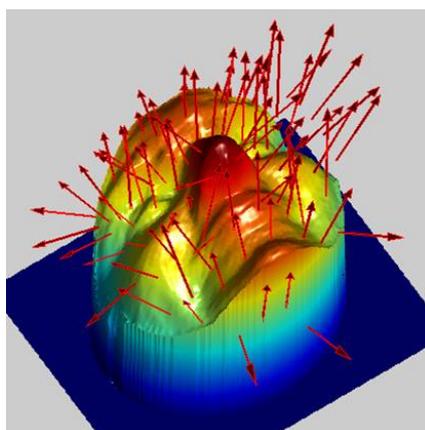
Таблица 5.10: Параметри на алгоритъма за разпознаване на 3D лица с множество ядра.

Параметър	Стойност
Размер на радиуса на сферата за извличане на лице	100мм
Размер на медианният филтър за точков шум	5×5
Размер на прозорец за нормали	5×5 елемента
Размер на прозорец на признаците LARK	7×7 елемента
Параметри на полиномното ядро $\{\eta, \zeta\}$	$[-10 : 10]$, стъпка 1 (експер.)
Параметър на гаусовото ядро κ	$[0.5 : 5]$, стъпка 0.5 (експер.)
Брой ядра в множеството ядра	$21 + 10 = 31$



Фигура 5.8: Оригинално 3D сканирано изображение от SHREC, и лицевият регион след предварителната обработка.

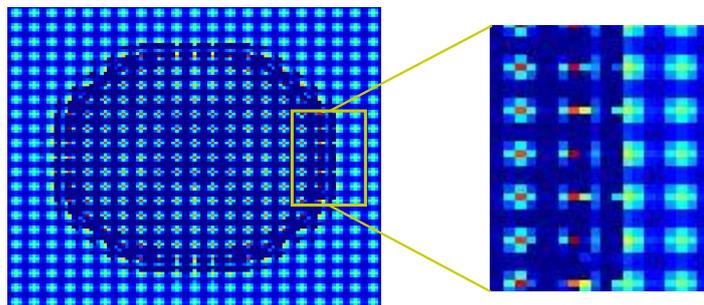
На Фиг. 5.9 са представени нормалите към повърхнината на 3D изображение на лицето, представено на Фиг. 5.8 (Около 60% от нормалите не са представени, за по-добра визуализация).



Фигура 5.9: Примерна лицева повърхнина със съответните ѝ нормални вектори към нея.

На Фиг. 5.10 е представена визуализация на признаците LARK, извлечени от прозорци без застъпване, от лицето представено на Фиг. 5.8. На увеличената част

от изображението, може да се забележи как признаците са адаптирани към локална част от повърхнината.



ФИГУРА 5.10: Визуализация на признаците LARK.

За определяне на точността на предложения алгоритъм се използва средния коефициент на разпознаване, който представлява средната стойности на разпознаването на всички индивиди от тестовата извадка.

Резултатите от тестовете на алгоритъма за двата проведени експеримента, както и сравнение с резултатите от работата в [182], са представени в Таблица 5.11. Алгоритъмът, който е избран за сравнение, използва същата база данни и същия начин за определяне на точността, като за сравнение са дадени и резултатите от два експеримента, представени в тази работа.

ТАБЛИЦА 5.11: Резултати от експериментите за определяне на ефективност на предложия алгоритъм и сравнението му с други.

Алгоритъм	Разпознаване[%]
Резултати от [182] (експ. 1)	80.8
Резултати от [182] (експ. 2)	82.2
Експеримент 1	92.73
Експеримент 2	94.90

Анализ на резултатите.

Резултатите от проведените експерименти показват, че предложият алгоритъм, реализиран с комбинация на допълващи се признаци и множество ядра, дава по-добри резултати от алгоритмите, докладвани в литературата, което потвърждава ефективността му. Друг извод, който може да се направи е, че броят на изображенията в обучаващата извадка влияе значително на ефективността на алгоритъма, като добавянето на дори едни изображение, увеличава ефективността на алгоритъма с близо 2%.

5.4 Основни резултати и изводи

Основните резултати от работата в тази глава са следните:

1. Реализирани са всички алгоритмите предложени в дисертацията, използвайки програмните езици C++ и MATLAB.

2. Проведени са експерименти с бази данни на лица, които се използват в научната литература.
3. Резултатите от проведените експерименти доказват ефективността на предложените в дисертацията алгоритми.

Изводите от работата по тази глава са следните:

1. Необходимо е провеждането на експерименти за определяне на параметрите на ядрата, с които параметри се постига максимална ефективност.
2. Реализираните алгоритми не изискват голям изчислителен ресурс и могат да бъдат използвани на стандартни работни станции.

5.5 Приноси към Глава 5

Приложните приноси са:

1. Програмна реализация и експериментални изследвания на алгоритъм за детектиране на лица, с комбинацията от детектор и класификатор, и последваща регистрация на лица, чрез очният базис. Този принос е към Глави 2,3 и 5, точки 2.2, 3.2, 3.3, 5.1.1, 5.2.1 и 5.2.2.
2. Програмна реализация и експериментални изследвания на алгоритъм за определяне на възраст в групи с Spectral Regression. Постигнат коефициент на разпознаване от 77.67%, използвайки стандартна база данни от лица с анотирани възрасти. Този принос е към Глава 2 и 5, точки 2.2 и 5.1.1.
3. Програмна реализация и експериментални изследвания на алгоритъм за разпознаване на изражения на лица, използващ разреден избор на log-Gabor признаци. Постигнат коефициент на разпознаване от алгоритъма 96.7%, подобряващ докладваните в литературата резултати, използвайки стандартна база данни от лица с анотирани изражения. Този принос е към Глава 3 и 5, точки 3.2 и 5.2.1.
4. Програмна реализация и експериментални изследвания на алгоритъм за разпознаване на изражения на лица, използващ извличане на геометрични признаци на лице и апроксимация на метода с ядрото. Постигнат коефициент на разпознаване от алгоритъма 53.6%, подобряващ докладваните в литературата, използвайки стандартна база данни от лица с анотирани изражения. Този принос е към Глави 3 и 5, точки 3.3 и 5.2.2.
5. Програмна реализация и експериментални изследвания на алгоритъм за разпознаване на 3D лица, използващ извличане на геометрични признаци и проекция в подпространство. Постигнат коефициент на разпознаване от 96.64%, използвайки стандартна база данни от лица. Този принос е към Глави 4 и 5, точки 4.2 и 5.3.1.

6. Програмна реализация и експериментални изследвания на алгоритъм за разпознаване на 3D лица, използващ допълващи се геометрични признаци и интегрирано в SVM множество ядра. Постигнат коефициент на разпознаване от 94.90%, подобряващ докладваните в литературата, използвайки стандартна база данни от лица. Този принос е към Глави 4 и 5. Този принос е към Глави 4 и 5, точки 4.3 и 5.3.2.

Описаните резултати от изследванията в Глава 5 са публикувани в [D1], [D2], [D3], [D4] и [D5].

Приноси в дисертационния труд

Научните приноси са:

1. Разработен е комбиниран метод, позволяващ работа с големи обеми от данни, базиран на геометрични признаци на лицето и обучение чрез множество ядра, апроксимирани със случайна извадка от косинусови базиси. Този принос е към Глава 3, точка 3.3.
2. Разработен е метод, базиран на допълващи се геометрични признаци и комбинацията им с множество ядра, използвани като резултантно ядро в класификатора SVM. Този принос е към Глава 4, точка 4.3.

Научно-приложните приноси са:

1. Разработен е подобряващ дискриминативността алгоритъм, за определяне на възраст в групи, базиран на намаляване на размерността, чрез реализирано по метода с ядрото вграждане на възрастово многообразие в подпространство. Този принос е към Глава 2, точка 2.2.
2. Разработени са два алгоритъма за разпознаване на изражения на лица, базирани на локални и геометрични признаци, последващ избор на признаци чрез разрежена регресия и апроксимация на метода с ядрото със случайна извадка от косинусови базиси. Избора на признаци и апроксимацията на метода с ядрото подобряват дискриминативността и позволяват работа с големи обеми от данни. Този принос е към Глава 3, точки 3.2 и 3.3.
3. Разработен е подобряващ дискриминативността алгоритъм за разпознаване на 3D лица, използващ геометрични признаци, базирани на кривината на формата и проекцията им в подпространство. Този принос е към Глава 4, точка 4.2.
4. Разработен е алгоритъм за разпознаване на 3D лица, базиран на допълващи се геометрични признаци и оптималната им комбинация чрез множество ядра. Допълващите се геометрични признаци увеличават информативността, а оптималната им комбинация увеличава дискриминативността на класификатора. Този принос е към Глава 4, точка 4.3.

Приложните приноси са:

1. Програмна реализация и експериментални изследвания на алгоритъм за детектиране на лица, с комбинацията от детектор и класификатор, и последваща регистрация на лица, чрез очният базис. Този принос е към Глави 2,3 и 5, точки 2.2, 3.2, 3.3, 5.1.1, 5.2.1 и 5.2.2.

2. Програмна реализация и експериментални изследвания на алгоритъм за определяне на възраст в групи с Spectral Regression. Постигнат коефициент на разпознаване от 77.67%, използвайки стандартна база данни от лица с аотирани възрасти. Този принос е към Глава 2 и 5, точки 2.2 и 5.1.1.
3. Програмна реализация и експериментални изследвания на алгоритъм за разпознаване на изражения на лица, използващ разреден избор на log-Gabor признаци. Постигнат коефициент на разпознаване от алгоритъма 96.7%, подобряващ докладваните в литературата резултати, използвайки стандартна база данни от лица с аотирани изражения. Този принос е към Глава 3 и 5, точки 3.2 и 5.2.1.
4. Програмна реализация и експериментални изследвания на алгоритъм за разпознаване на изражения на лица, използващ извличане на геометрични признаци на лице и апроксимация на метода с ядрото. Постигнат коефициент на разпознаване от алгоритъма 53.6%, подобряващ докладваните в литературата, използвайки стандартна база данни от лица с аотирани изражения. Този принос е към Глави 3 и 5, точки 3.3 и 5.2.2.
5. Програмна реализация и експериментални изследвания на алгоритъм за разпознаване на 3D лица, използващ извличане на геометрични признаци и проекция в подпространство. Постигнат коефициент на разпознаване от 96.64%, използвайки стандартна база данни от лица. Този принос е към Глави 4 и 5, точки 4.2 и 5.3.1.
6. Програмна реализация и експериментални изследвания на алгоритъм за разпознаване на 3D лица, използващ допълващи се геометрични признаци и интегрирано в SVM множество ядра. Постигнат коефициент на разпознаване от 94.90%, подобряващ докладваните в литературата, използвайки стандартна база данни от лица. Този принос е към Глави 4 и 5. Този принос е към Глави 4 и 5, точки 4.3 и 5.3.2.

Авторски публикации по дисертационните изследвания

- [D1] **K. Tonchev**, I. Paliy , O. Boumbarov , S. Sokolov, "Human age-group classification of facial images with subspace projection and support vector machines." *Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems*. Vol. 1. IEEE, 2011. DOI: 10.1109/IDAACS.2011.6072792
- [D2] **K. Tonchev**, N. Neshov, A. Manolova, V. Poulkov, "Expression recognition using sparse selection of log-Gabor facial features. In 2017 *Fourth International Conference on Mathematics and Computers in Sciences and in Industry (MCSI)*, (pp. 11-15). IEEE.
- [D3] **K. Tonchev**, N. Neshov, A. Manolova, T. Sechkova, "Applications of large scale kernel machines for real world human mood estimation." *International Journal of Signal Processing, Volume*, vol. 2, 2017, ISSN: 2367-8984
- [D4] **K. Tonchev**, A. Manolova, I. Paliy, "Comparative analysis of 3d face recognition algorithms using range image and curvature-based representations." In 2013 *IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS)* (Vol. 1, pp. 394-398). IEEE. DOI: 10.1109/IDAACS.2013.6662714
- [D5] **K. Tonchev**, "3D Face Recognition with Multiple Kernel Learning." *WSEAS International Conference. Proceedings. Recent Advances in Computer Engineering Series*. No. 7. WSEAS, 2012. ISBN: 978-1-61804-134-0

Участие в научно-изследователски проекти по време на докторантурата

Договор, финансиран от Европейската комисия

1. Grant Nr. SIP-2008-TP-131801 “Forensic Image and Video Examination Support“ (FIVES) of the European Commission, “Safer Internet Plus” Programme, 2009-2011 г., участник, приключил успешно.

Договори към Фонд “Научни изследвания” на МОН

2. Дог. ДОО2-41/2008 „Биометрична идентификация в системи за видеонаблюдение”, Билатерален договор между България и Украйна, Финансиран от Фонд Научни Изследвания при МОН, 2008-2011, участник, приключил успешно.
3. Дог. ДНТС/Словения 01/8 „3D разпознаване на лица: методи и алгоритми с повишена надеждност и бързодействие“, Билатерален договор между България и Словения, Финансиран от Фонд Научни Изследвания при МОН, 2010-2014, участник, приключил успешно.
4. Дог. ДФНИ И 02/1 „Интелигентен интерфейс човек-компютър за асистирани медицински системи, подобряващи качеството на живот на хора с двигателни увреждания“, Финансиран от Фонд Научни Изследвания при МОН, 2014-2019, участник, приключил успешно.

Списък на цитиранията

Цитирания на публикация D1:

1. Gnana Rajesh, D., and M. Punithavalli. "An Efficient Age Group Classification System using Selected Undecimated Wavelet Coefficients of Fingerprints." *International Journal of Applied Engineering Research*, vol. 9, no. 22, 2014, pp. 14303-14318.
2. Priyadarshni, V., et al. "Human Age Classification System using K-NN Classifier vol. 1075, 2019. doi:10.1007/978-981-15-0108-1_28.
3. Yilmaz, A. G., and V. Nabiyev. "A Novel Age Classification Method using Morph-Based Models." *International Arab Journal of Information Technology*, vol. 16, no. 4, 2019, pp. 677-685.

Цитирания на публикация D3:

4. Han, Z., H. Huang, and J. Wang. Convolutional Neural Network Based Expression Classification with Face Alignment, 2018. doi:10.1109/ICCSP.2018.8572342.

Цитирания на публикация D4:

5. Hauenstein, J. D., and T. S. Newman. "Curvature Determination in Range Images: New Methods and Comparison Study." *Multimedia Tools and Applications*, vol. 78, no. 7, 2019, pp. 9247-9273.
6. Maneli, M. A., and O. E. Isafiade. "3D Forensic Crime Scene Reconstruction Involving Immersive Technology: A Systematic Literature Review." *IEEE Access*, vol. 10, 2022, pp. 88821-88857, doi:10.1109/ACCESS.2022.3199437.
7. Tang, Y., et al. "3D Face Recognition with Asymptotic Cones Based Principal Curvatures 2015. doi:10.1109/ICB.2015.7139111.
8. Zhang, K., S. Qiao, and K. Gao. "A New Point Cloud Reconstruction Algorithm Based-on Geometrical Features 2016. doi:10.1109/ICMIC.2015.7409387.
9. Zhang, K., S. Qiao, and K. Gao. "Surface Reconstruction Algorithm Based on Local Data Features." *International Journal of Modelling, Identification and Control*, vol. 30, no. 3, 2018, pp. 197-205, doi:10.1504/IJMIC.2018.095336.

Заклучение и бъдеща работа

Предложените в този труд алгоритми за анализ на изображения на лица демонстрират ефективност и ефикасност които са достатъчно добри да бъдат използвани на практика и да улеснят взаимодействието между човек и компютър. Основният компонент, линейните статистически преобразувания и реализацията им по метода с ядрото, доказват възможността да бъдат успешно използвани в разнообразни приложения, защото предоставят висок капацитет на обучението или бързодействие което е оптимално спрямо желаната ефективност. Макар метода с ядрото да не е ефективен с нарастването на данните, неговите приближения позволяват да се заобиколи този проблем и към момента те са тема на активна изследователска дейност, защото предоставят възможност да се интерпретира поведението им, за разлика от невронните мрежи. Нещо повече, през изминалата година има ново направление на изследователска дейност в основата на което стои Невронното Тангенциално Ядро (Neural Tangent Kernel - НТК) и което се явява мост между съвременните невронни мрежи и метода с ядрото.

Бъдещите насоки на работа са свързани именно с използването на НТК и интеграцията му като част от невронни мрежи реализирани по механизма с „внимание“, целящи интерпретация на поведението им. Също така ще бъдат изследвани нови методи, които използват приближение на матрицата на Грам и позволява работа с големи обеми от данни. Изследването на разгледаните в този труд приложения ще бъде задълбочено като се наблегне на анализ в тримерното пространство.

Исползвана литература

- [1] R. Geirhos, C. R. Temme, J. Rauber, H. H. Schütt, M. Bethge, and F. A. Wichmann, „Generalisation in humans and deep neural networks“, *Advances in neural information processing systems*, 31, no., 2018.
- [2] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, „Classifying facial actions“, *IEEE Transactions on pattern analysis and machine intelligence*, 21, no., pp. 974–989, 1999.
- [3] M. Pantic and L. J. M. Rothkrantz, „Automatic analysis of facial expressions: The state of the art“, *IEEE Transactions on pattern analysis and machine intelligence*, 22, no., pp. 1424–1445, 2000.
- [4] M. S. Nixon, P. L. Correia, K. Nasrollahi, T. B. Moeslund, A. Hadid, and M. Tistarelli, „On soft biometrics“, *Pattern Recognition Letters*, 68, no., pp. 218–230, 2015.
- [5] A. van Dam, „Beyond wimp“, *IEEE Computer Graphics and Applications*, 20, no., pp. 50–51, 2000.
- [6] A. G.T.S. H. Nicu Sebe Ira Cohen, *Application: Facial Expression Recognition. In: Machine Learning in Computer Vision, Computational Imaging and Vision, vol 29*. Springer, Dordrecht, 2005.
- [7] P. Ekman and W. V. Friesen, „Facial action coding system“, *Environmental Psychology & Nonverbal Behavior*, no., 1978.
- [8] J. WESTERMEYER, „A social interactional theory of emotions“, *American Journal of Psychiatry*, 136, no., pp. 870–870, 1979.
- [9] C. Darwin, *The expression of the emotions in man and animals by Charles Darwin*. John Murray, 1872.
- [10] W. A. Freiwald, D. Y. Tsao, and M. S. Livingstone, „A face feature space in the macaque temporal lobe“, *Nature neuroscience*, 12, no., p. 1187, 2009.
- [11] G. Rhodes and J. Haxby, *Oxford handbook of face perception*. Oxford University Press, 2011.
- [12] G. W. Cottrell, M. N. Dailey, C. Padgett, and R. Adolphs, „Is all face processing holistic? The view from UCSD“, *Computational, geometric, and process perspectives on facial cognition*, no., pp. 347–396, 2001.
- [13] D. C. Neth, „Facial configuration and the perception of facial expression“, PhD thesis, The Ohio State University, 2007.
- [14] M. White, „Parts and wholes in expression recognition“, *Cognition & Emotion*, 14, no., pp. 39–60, 2000.

- [15] L. Zhang and G. W. Cottrell, „When holistic processing is not enough: Local features save the day“, in *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 26, 2004.
- [16] N. Bourbaki, *Elements of mathematics: Theory of sets*. Hermann (original), Addison-Wesley (translation), 1968.
- [17] W. Yang, H. Du, Z. Liew, W. Y. B. Lim, Z. Xiong, D. Niyato, X. Chi, X. S. Shen, and C. Miao, „Semantic communications for 6G future internet: Fundamentals, applications, and challenges“, *arXiv preprint arXiv:2207.00427*, no., 2022.
- [18] S. Pittner and S. V. Kamarthi, „Feature extraction from wavelet coefficients for pattern recognition tasks“, *IEEE Transactions on pattern analysis and machine intelligence*, 21, no., pp. 83–88, 1999.
- [19] S. Kakarwal and R. Deshmukh, „Wavelet transform based feature extraction for face recognition“, *IJCSA Issue-I June*, 9740767, no., 2010.
- [20] G. Guo, G. Mu, Y. Fu, and T. S. Huang, „Human age estimation using bio-inspired features“, in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, IEEE, 2009, pp. 112–119.
- [21] M. S. Bartlett, G. Littlewort, M. G. Frank, C. Lainscsek, I. R. Fasel, J. R. Movellan, *et al.*, „Automatic recognition of facial actions in spontaneous expressions.“, *J. Multim.*, 1, no., pp. 22–35, 2006.
- [22] N. Rose, „Facial expression classification using gabor and log-gabor filters“, in *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, IEEE, 2006, pp. 346–350.
- [23] X. Gao, F. Sattar, and R. Venkateswarlu, „Multiscale corner detection of gray level images based on log-gabor wavelet transform“, *IEEE Transactions on Circuits and Systems for Video Technology*, 17, no., pp. 868–875, 2007.
- [24] T. Ahonen, A. Hadid, and M. Pietikainen, „Face description with local binary patterns: Application to face recognition“, *IEEE transactions on pattern analysis and machine intelligence*, 28, no., pp. 2037–2041, 2006.
- [25] X. Tan and B. Triggs, „Enhanced local texture feature sets for face recognition under difficult lighting conditions“, *IEEE transactions on image processing*, 19, no., pp. 1635–1650, 2010.
- [26] M. Dahmane and J. Meunier, „Emotion recognition using dynamic grid-based HoG features“, in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, IEEE, 2011, pp. 884–888.
- [27] M. Bicego, A. Lagorio, E. Grosso, and M. Tistarelli, „On the use of SIFT features for face authentication“, in *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*, IEEE, 2006, pp. 35–35.
- [28] G. Du, F. Su, and A. Cai, „Face recognition using SURF features“, in *MIPPR 2009: Pattern Recognition and Computer Vision*, International Society for Optics and Photonics, vol. 7496, 2009, p. 749 628.
- [29] V. Ojansivu and J. Heikkilä, „Blur insensitive texture classification using local phase quantization“, in *International conference on image and signal processing*, Springer, 2008, pp. 236–243.

- [30] Y. Zhu, F. De la Torre, J. F. Cohn, and Y.-J. Zhang, „Dynamic cascades with bidirectional bootstrapping for action unit detection in spontaneous facial behavior“, *IEEE transactions on affective computing*, 2, no., pp. 79–91, 2011.
- [31] E. Sariyanidi, H. Gunes, M. Gökmen, and A. Cavallaro, „Local Zernike Moment Representation for Facial Affect Recognition.“, in *BMVC*, vol. 2, 2013, p. 3.
- [32] K. Sikka, T. Wu, J. Susskind, and M. Bartlett, „Exploring bag of words architectures in the facial expression domain“, in *European Conference on Computer Vision*, Springer, 2012, pp. 250–259.
- [33] S. F. Cotter, „Sparse representation for accurate classification of corrupted and occluded facial expressions“, in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, 2010, pp. 838–841.
- [34] E. Tola, V. Lepetit, and P. Fua, „Daisy: An efficient dense descriptor applied to wide-baseline stereo“, *IEEE transactions on pattern analysis and machine intelligence*, 32, no., pp. 815–830, 2009.
- [35] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikäinen, X. Chen, and W. Gao, „WLD: A robust local image descriptor“, *IEEE transactions on pattern analysis and machine intelligence*, 32, no., pp. 1705–1720, 2009.
- [36] E. Shechtman and M. Irani, „Matching local self-similarities across images and videos“, in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2007, pp. 1–8.
- [37] M. A. Turk and A. P. Pentland, „Face recognition using eigenfaces“, in *Proceedings. 1991 IEEE computer society conference on computer vision and pattern recognition*, IEEE Computer Society, 1991, pp. 586–587.
- [38] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, „Eigenfaces vs. fisherfaces: Recognition using class specific linear projection“, *IEEE Transactions on pattern analysis and machine intelligence*, 19, no., pp. 711–720, 1997.
- [39] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, „Face recognition by independent component analysis“, *IEEE Transactions on neural networks*, 13, no., pp. 1450–1464, 2002.
- [40] S. Nikitidis, A. Tefas, N. Nikolaidis, and I. Pitas, „Subclass discriminant nonnegative matrix factorization for facial image analysis“, *Pattern Recognition*, 45, no., pp. 4080–4091, 2012.
- [41] K. I. Kim, K. Jung, and H. J. Kim, „Face recognition using kernel principal component analysis“, *IEEE signal processing letters*, 9, no., pp. 40–42, 2002.
- [42] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.-R. Mullers, „Fisher discriminant analysis with kernels“, in *Neural networks for signal processing IX: Proceedings of the 1999 IEEE signal processing society workshop (cat. no. 98th8468)*, Ieee, 1999, pp. 41–48.
- [43] G. Baudat and F. Anouar, „Generalized discriminant analysis using a kernel approach“, *Neural computation*, 12, no., pp. 2385–2404, 2000.
- [44] S. T. Roweis and L. K. Saul, „Nonlinear dimensionality reduction by locally linear embedding“, *science*, 290, no., pp. 2323–2326, 2000.

- [45] M. Belkin and P. Niyogi, „Laplacian eigenmaps for dimensionality reduction and data representation“, *Neural computation*, 15, no., pp. 1373–1396, 2003.
- [46] X. He and P. Niyogi, „Locality preserving projections“, *Advances in neural information processing systems*, 16, no., pp. 153–160, 2004.
- [47] D. Cai, X. He, J. Han, and H.-J. Zhang, „Orthogonal laplacianfaces for face recognition“, *IEEE transactions on image processing*, 15, no., pp. 3608–3614, 2006.
- [48] H.-T. Chen, H.-W. Chang, and T.-L. Liu, „Local discriminant embedding and its variants“, in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, IEEE, vol. 2, 2005, pp. 846–853.
- [49] D. Cai, *Spectral regression: A regression framework for efficient regularized subspace learning*. ProQuest, 2009.
- [50] Y. Pang, L. Zhang, Z. Liu, N. Yu, and H. Li, „Neighborhood preserving projections (NPP): A novel linear dimension reduction method“, in *International conference on intelligent computing*, Springer, 2005, pp. 117–125.
- [51] Y. Fu, M. Liu, and T. S. Huang, „Conformal embedding analysis with local graph modeling on the unit hypersphere“, in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2007, pp. 1–6.
- [52] A. Krizhevsky, I. Sutskever, and G. E. Hinton, „Imagenet classification with deep convolutional neural networks“, *Advances in neural information processing systems*, 25, no., pp. 1097–1105, 2012.
- [53] K. Simonyan and A. Zisserman, „Very deep convolutional networks for large-scale image recognition“, *arXiv preprint arXiv:1409.1556*, no., 2014.
- [54] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, „Going deeper with convolutions“, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
- [55] K. He, X. Zhang, S. Ren, and J. Sun, „Deep residual learning for image recognition“, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [56] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, „You only look once: Unified, real-time object detection“, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
- [57] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, *et al.*, „An image is worth 16x16 words: Transformers for image recognition at scale“, *arXiv preprint arXiv:2010.11929*, no., 2020.
- [58] L. Yuan, Q. Hou, Z. Jiang, J. Feng, and S. Yan, „Volo: Vision outlooker for visual recognition“, *arXiv preprint arXiv:2106.13112*, no., 2021.
- [59] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, F. Wei, and B. Guo, „Swin Transformer V2: Scaling Up Capacity and Resolution“, in *International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

- [60] N. Li, Y. Chen, W. Li, Z. Ding, and D. Zhao, „BViT: Broad Attention based Vision Transformer“, *arXiv preprint arXiv:2202.06268*, no., 2022.
- [61] M. Vaishnav, T. Fel, I. F. Rodriguez, and T. Serre, „Conviformers: Convolutionally guided Vision Transformer“, *arXiv preprint arXiv:2208.08900*, no., 2022.
- [62] Y. Li, H. Mao, R. Girshick, and K. He, „Exploring plain vision transformer backbones for object detection“, *arXiv preprint arXiv:2203.16527*, no., 2022.
- [63] A. Ranjan, T. Bolkart, S. Sanyal, and M. J. Black, „Generating 3D faces using convolutional mesh autoencoders“, in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 704–720.
- [64] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, „Dynamic graph cnn for learning on point clouds“, *Acm Transactions On Graphics (tog)*, 38, no., pp. 1–12, 2019.
- [65] R. Hanocka, A. Hertz, N. Fish, R. Giryes, S. Fleishman, and D. Cohen-Or, „Meshcnn: a network with an edge“, *ACM Transactions on Graphics (TOG)*, 38, no., pp. 1–12, 2019.
- [66] A. Tewari, F. Bernard, P. Garrido, G. Bharaj, M. Elgharib, H.-P. Seidel, P. Pérez, M. Zollhofer, and C. Theobalt, „Fml: Face model learning from videos“, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 812–10 822.
- [67] N. Kolotouros, G. Pavlakos, and K. Daniilidis, „Convolutional mesh regression for single-image human shape reconstruction“, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4501–4510.
- [68] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, „Deep learning enabled semantic communication systems“, *IEEE Transactions on Signal Processing*, 69, no., pp. 2663–2675, 2021.
- [69] D. Huang, F. Gao, X. Tao, Q. Du, and J. Lu, „Towards Semantic Communications: Deep Learning-Based Image Semantic Coding“, *arXiv preprint arXiv:2208.04094*, no., 2022.
- [70] O. Ronneberger, P. Fischer, and T. Brox, „U-net: Convolutional networks for biomedical image segmentation“, in *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.
- [71] B. Graham, M. Engelcke, and L. Van Der Maaten, „3d semantic segmentation with submanifold sparse convolutional networks“, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 9224–9232.
- [72] R. Strudel, R. Garcia, I. Laptev, and C. Schmid, „Segmenter: Transformer for semantic segmentation“, in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 7262–7272.
- [73] K. Fukunaga, *Introduction to statistical pattern recognition*. Elsevier, 2013.
- [74] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik, „Support vector regression machines“, *Advances in neural information processing systems*, 9, no., 1996.
- [75] W. M. Boothby and W. M. Boothby, *An introduction to differentiable manifolds and Riemannian geometry, Revised*. Gulf Professional Publishing, 2003, vol. 120.

- [76] C. Esteves, C. Allen-Blanchette, A. Makadia, and K. Daniilidis, „Learning so (3) equivariant representations with spherical cnns“, in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 52–68.
- [77] L. M. Zouhal and T. Denoex, „An evidence-theoretic k-NN rule with parameter optimization“, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 28, no., pp. 263–271, 1998.
- [78] H. Zhu, K. Sun, and P. Koniusz, „Contrastive laplacian eigenmaps“, *Advances in Neural Information Processing Systems*, 34, no., pp. 5682–5695, 2021.
- [79] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, „Graph embedding and extensions: A general framework for dimensionality reduction“, *IEEE transactions on pattern analysis and machine intelligence*, 29, no., pp. 40–51, 2006.
- [80] J. Shawe-Taylor, N. Cristianini, *et al.*, *Kernel methods for pattern analysis*. Cambridge university press, 2004.
- [81] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, „Face recognition using kernel direct discriminant analysis algorithms“, *IEEE Transactions on Neural Networks*, 14, no., pp. 117–126, 2003.
- [82] Y. H. Kwon and N. da Vitoria Lobo, „Age classification from facial images“, *Computer vision and image understanding*, 74, no., pp. 1–21, 1999.
- [83] A. Báez-Suárez, C. Nikou, J. A. Nolasco-Flores, and I. A. Kakadiaris, „Age Classification from Facial Images: Is Frontalization Necessary?“, in *International Symposium on Visual Computing*, Springer, 2016, pp. 769–778.
- [84] G. Levi and T. Hassner, „Age and gender classification using convolutional neural networks“, in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2015, pp. 34–42.
- [85] R. Angulu, J. R. Tapamo, and A. O. Adewumi, „Age estimation via face images: a survey“, *EURASIP Journal on Image and Video Processing*, 2018, no., pp. 1–35, 2018.
- [86] X. Geng, Z.-H. Zhou, and K. Smith-Miles, „Automatic age estimation based on facial aging patterns“, *IEEE Transactions on pattern analysis and machine intelligence*, 29, no., pp. 2234–2240, 2007.
- [87] Y. Fu and T. S. Huang, „Human age estimation with regression on discriminative aging manifold“, *IEEE Transactions on Multimedia*, 10, no., pp. 578–584, 2008.
- [88] S. Yan, H. Wang, X. Tang, J. Liu, and T. S. Huang, „Regression from uncertain labels and its applications to soft biometrics“, *IEEE Transactions on Information Forensics and Security*, 3, no., pp. 698–708, 2008.
- [89] L. G. Farkas, *Anthropometry of the Head and Face*. Lippincott Williams & Wilkins, 1994.
- [90] T. R. Alley, *Social and applied aspects of perceiving faces*. Psychology Press, 2013.
- [91] N. Ramanathan and R. Chellappa, „Modeling age progression in young faces“, in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, IEEE, vol. 1, 2006, pp. 387–394.

- [92] A. Gunay and V. V. Nabiyev, „Automatic detection of anthropometric features from facial images“, in *2007 IEEE 15th Signal Processing and Communications Applications*, IEEE, 2007, pp. 1–4.
- [93] X. Geng, Z.-H. Zhou, Y. Zhang, G. Li, and H. Dai, „Learning from facial aging patterns for automatic age estimation“, in *Proceedings of the 14th ACM international conference on Multimedia*, 2006, pp. 307–316.
- [94] G. J. Edwards, A. Lanitis, C. J. Taylor, and T. F. Cootes, „Statistical models of face images—Improving specificity“, *Image and Vision Computing*, 16, no., pp. 203–211, 1998.
- [95] G. Panis and A. Lanitis, „An overview of research activities in facial age estimation using the FG-NET aging database“, in *European Conference on Computer Vision*, Springer, 2014, pp. 737–750.
- [96] S. Yan, H. Wang, X. Tang, J. Liu, and T. S. Huang, „Regression from uncertain labels and its applications to soft biometrics“, *IEEE Transactions on Information Forensics and Security*, 3, no., pp. 698–708, 2008.
- [97] S. Yan, X. Zhou, M. Liu, M. Hasegawa-Johnson, and T. S. Huang, „Regression from patch-kernel“, in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2008, pp. 1–8.
- [98] C. N. Duong, K. G. Quach, K. Luu, H. B. Le, and K. Ricanek, „Fine tuning age estimation with global and local facial features“, in *IEEE International Conference on Acoustics, Speech and Signal Processing, Prague, Czech Republic*, 2011, pp. 2032–2035.
- [99] S. Yan, M. Liu, and T. S. Huang, „Extracting age information from local spatially flexible patches“, in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, 2008, pp. 737–740.
- [100] K. Luu, T. Dai Bui, and C. Y. Suen, „Kernel spectral regression of perceived age from hybrid facial features“, in *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, IEEE, 2011, pp. 1–6.
- [101] C. Chen, W. Yang, Y. Wang, K. Ricanek, and K. Luu, „Facial feature fusion and model selection for age estimation“, in *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, IEEE, 2011, pp. 200–205.
- [102] M. Y. El Dib and M. El-Saban, „Human age estimation using enhanced bio-inspired features (EBIF)“, in *2010 IEEE International Conference on Image Processing*, IEEE, 2010, pp. 1589–1592.
- [103] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, „Image-based human age estimation by manifold learning and locally adjusted robust regression“, *IEEE Transactions on Image Processing*, 17, no., pp. 1178–1188, 2008.
- [104] P. Turaga, S. Biswas, and R. Chellappa, „The role of geometry in age estimation“, in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, 2010, pp. 946–949.
- [105] C. Cortes and V. Vapnik, „Support-vector networks“, *Machine learning*, 20, no., pp. 273–297, 1995.

- [106] P. Viola and M. J. Jones, „Robust real-time face detection“, *International journal of computer vision*, 57, no., pp. 137–154, 2004.
- [107] R. S. Stanković and B. J. Falkowski, „The Haar wavelet transform: its status and achievements“, *Computers & Electrical Engineering*, 29, no., pp. 25–44, 2003.
- [108] R. E. Schapire, „Explaining adaboost“, in *Empirical inference*, Springer, 2013, pp. 37–52.
- [109] Y. Freund, R. E. Schapire, *et al.*, „Experiments with a new boosting algorithm“, in *icml*, Citeseer, vol. 96, 1996, pp. 148–156.
- [110] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, „Gradient-based learning applied to document recognition“, *Proceedings of the IEEE*, 86, no., pp. 2278–2324, 1998.
- [111] J. Zhou, L. Jiang, Z. Ji, and L. Shen, „Haar-like features based eye detection algorithm and its implementation on TI TMS320DM6446 platform“, in *2009 IEEE International Workshop on Imaging Systems and Techniques*, IEEE, 2009, pp. 89–93.
- [112] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*, 4. Springer, 2006, vol. 4.
- [113] D. Cai, X. He, and J. Han, „Spectral regression for efficient regularized subspace learning“, in *2007 IEEE 11th international conference on computer vision*, IEEE, 2007, pp. 1–8.
- [114] A. Ng, M. Jordan, and Y. Weiss, „On spectral clustering: Analysis and an algorithm“, *Advances in neural information processing systems*, 14, no., 2001.
- [115] P. Ekman and W. V. Friesen, „Constants across cultures in the face and emotion.“, *Journal of personality and social psychology*, 17, no., p. 124, 1971.
- [116] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, „Coding facial expressions with gabor wavelets“, in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, IEEE, 1998, pp. 200–205.
- [117] V. J. Mistry and M. M. Goyani, „A literature survey on facial expression recognition using global features“, *Int. J. Eng. Adv. Technol*, 2, no., 653–657, 2013.
- [118] Y. Tian, T. Kanade, and J. F. Cohn, „Facial expression recognition“, in *Handbook of face recognition*, Springer, 2011, pp. 487–519.
- [119] A. Martinez and R. Benavente, „The ar face database: Cvc technical report, 24“, no., 1998.
- [120] C. Shan, S. Gong, and P. W. McOwan, „Facial expression recognition based on local binary patterns: A comprehensive study“, *Image and Vision Computing*, 27, no., pp. 803–816, 2009.
- [121] J. Ou, X.-B. Bai, Y. Pei, L. Ma, and W. Liu, „Automatic facial expression recognition using Gabor filter and expression analysis“, in *Computer Modeling and Simulation, 2010. ICCMS'10. Second International Conference on*, IEEE, vol. 2, 2010, pp. 215–218.

- [122] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, „The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression“, in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, IEEE, 2010, pp. 94–101.
- [123] R. S. El-Sayed and A. El Kholy, „Robust Facial Expression Recognition via Sparse Representation and Multiple Gabor filters“, no.,
- [124] S. M. Lajevardi and Z. M. Hussain, „Feature selection for facial expression recognition based on optimization algorithm“, in *Nonlinear Dynamics and Synchronization, 2009. INDS'09. 2nd International Workshop on*, IEEE, 2009, pp. 182–185.
- [125] S. M. Lajevardi and Z. M. Hussain, „Feature extraction for facial expression recognition based on hybrid face regions“, *Advances in Electrical and Computer Engineering*, 9, no., pp. 63–67, 2009.
- [126] N. Mehta and S. Jadhav, „Facial Emotion recognition using Log Gabor filter and PCA“, in *Computing Communication Control and automation (ICCUBEA), 2016 International Conference on*, IEEE, 2016, pp. 1–5.
- [127] C. A. Corneanu, M. O. Simón, J. F. Cohn, and S. E. Guerrero, „Survey on rgb, 3d, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications“, *IEEE transactions on pattern analysis and machine intelligence*, 38, no., pp. 1548–1568, 2016.
- [128] A. Dhall, O. Ramana Murthy, R. Goecke, J. Joshi, and T. Gedeon, „Video and image based emotion recognition challenges in the wild: Emotiw 2015“, in *Proceedings of the 2015 ACM on international conference on multimodal interaction*, 2015, pp. 423–426.
- [129] S. Ebrahimi Kahou, V. Michalski, K. Konda, R. Memisevic, and C. Pal, „Recurrent neural networks for emotion recognition in video“, in *Proceedings of the 2015 ACM on international conference on multimodal interaction*, 2015, pp. 467–474.
- [130] G. Levi and T. Hassner, „Emotion recognition in the wild via convolutional neural networks and mapped binary patterns“, in *Proceedings of the 2015 ACM on international conference on multimodal interaction*, 2015, pp. 503–510.
- [131] I. Borg and P. J. Groenen, *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media, 2005.
- [132] X. Zhitao, G. Chengming, Y. Ming, and L. Qiang, „Research on log Gabor wavelet and its application in image edge detection“, in *6th International Conference on Signal Processing, 2002.*, IEEE, vol. 1, 2002, pp. 592–595.
- [133] L. Shen and L. Bai, „A review on Gabor wavelets for face recognition“, *Pattern analysis and applications*, 9, no., pp. 273–292, 2006.
- [134] D. J. Field, „Relations between the statistics of natural images and the response properties of cortical cells“, *JOSA A*, 4, no., pp. 2379–2394, 1987.
- [135] D. Cai, C. Zhang, and X. He, „Unsupervised feature selection for multi-cluster data“, in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2010, pp. 333–342.

- [136] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, „Least angle regression“, *The Annals of statistics*, 32, no., pp. 407–499, 2004.
- [137] X. Xiong and F. De la Torre, „Supervised descent method and its applications to face alignment“, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 532–539.
- [138] T. F. Cootes and C. J. Taylor, „Active shape models—‘smart snakes’“, in *BMVC92*, Springer, 1992, pp. 266–275.
- [139] T. F. Cootes, G. J. Edwards, and C. J. Taylor, „Active appearance models“, *IEEE Transactions on pattern analysis and machine intelligence*, 23, no., pp. 681–685, 2001.
- [140] T. F. Cootes and C. J. Taylor, „Constrained active appearance models“, in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, IEEE, vol. 1, 2001, pp. 748–754.
- [141] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar, „Localizing parts of faces using a consensus of exemplars“, *IEEE transactions on pattern analysis and machine intelligence*, 35, no., pp. 2930–2940, 2013.
- [142] H. Yang and I. Patras, „Sieving regression forest votes for facial feature detection in the wild“, in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1936–1943.
- [143] L. Bottou, „Stochastic gradient descent tricks“, in *Neural networks: Tricks of the trade*, Springer, 2012, pp. 421–436.
- [144] Z. Lu, A. May, K. Liu, A. B. Garakani, D. Guo, A. Bellet, L. Fan, M. Collins, B. Kingsbury, M. Picheny, *et al.*, „How to scale up kernel methods to be as good as deep neural nets“, *arXiv preprint arXiv:1411.4000*, no., 2014.
- [145] A. Rahimi and B. Recht, „Weighted sums of random kitchen sinks: Replacing minimization with randomization in learning“, in *Advances in neural information processing systems*, 2009, pp. 1313–1320.
- [146] N. Roux, M. Schmidt, and F. Bach, „A stochastic gradient method with an exponential convergence rate for finite training sets“, *Advances in neural information processing systems*, 25, no., 2012.
- [147] M. Gönen and E. Alpaydm, „Multiple kernel learning algorithms“, *The Journal of Machine Learning Research*, 12, no., pp. 2211–2268, 2011.
- [148] Z. Eng, Y. Yick, Y. Guo, H. Xu, M. Reiner, T. Cham, and S. Chen, „3D faces are recognized more accurately and faster than 2D faces, but with similar inversion effects“, *Vision research*, 138, no., pp. 78–85, 2017.
- [149] C.-W. Wang and C.-C. Peng, „3D face point cloud reconstruction and recognition using depth sensor“, *Sensors*, 21, no., p. 2587, 2021.
- [150] A. Scheenstra, A. Ruifrok, and R. C. Veltkamp, „A survey of 3D face recognition methods“, in *International Conference on Audio-and Video-based Biometric Person Authentication*, Springer, 2005, pp. 891–899.
- [151] M. Li, B. Huang, and G. Tian, „A comprehensive survey on 3D face recognition methods“, *Engineering Applications of Artificial Intelligence*, 110, no., p. 104669, 2022.

- [152] S. Soltanpour, B. Boufama, and Q. J. Wu, „A survey of local feature methods for 3D face recognition“, *Pattern Recognition*, 72, no., pp. 391–406, 2017.
- [153] A. S. Mian, M. Bennamoun, and R. Owens, „Keypoint detection and local feature matching for textured 3D face recognition“, *International Journal of Computer Vision*, 79, no., pp. 1–12, 2008.
- [154] N. Bayramoglu and A. A. Alatan, „Shape index SIFT: Range image recognition using local features“, in *2010 20th International Conference on Pattern Recognition*, IEEE, 2010, pp. 352–355.
- [155] H. Zhou, A. Mian, L. Wei, D. Creighton, M. Hossny, and S. Nahavandi, „Recent advances on singlemodal and multimodal face recognition: a survey“, *IEEE Transactions on Human-Machine Systems*, 44, no., pp. 701–716, 2014.
- [156] Y. Lei, M. Bennamoun, M. Hayat, and Y. Guo, „An efficient 3D face recognition approach using local geometrical signatures“, *Pattern Recognition*, 47, no., pp. 509–524, 2014.
- [157] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, „Overview of the face recognition grand challenge“, in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, IEEE, vol. 1, 2005, pp. 947–954.
- [158] F. B. Ter Haar, M. Daoudi, and R. C. Veltkamp, „SHape REtrieval contest 2008: 3D face scans“, in *IEEE International Conference on Shape Modeling and Applications*, IEEE Computer Society, 2008, pp. 225–226.
- [159] H. Li, D. Huang, P. Lemaire, J.-M. Morvan, and L. Chen, „Expression robust 3D face recognition via mesh-based histograms of multiple order surface differential quantities“, in *2011 18th IEEE International Conference on Image Processing*, IEEE, 2011, pp. 3053–3056.
- [160] A. Savran, N. Alyüz, H. Dibeklioglu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, „Bosphorus database for 3D face analysis“, in *European workshop on biometrics and identity management*, Springer, 2008, pp. 47–56.
- [161] S. Berretti, N. Werghi, A. Del Bimbo, and P. Pala, „Selecting stable keypoints and local descriptors for person identification using 3D face scans“, *The Visual Computer*, 30, no., pp. 1275–1292, 2014.
- [162] S. Berretti, A. Del Bimbo, and P. Pala, „3D face recognition using isogeodesic stripes“, *IEEE transactions on pattern analysis and machine intelligence*, 32, no., pp. 2162–2177, 2010.
- [163] M. Mayo and E. Zhang, „3D face recognition using multiview keypoint matching“, in *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, IEEE, 2009, pp. 290–295.
- [164] A. Moreno, „GavabDB: a 3D face database“, in *Proc. 2nd COST275 Workshop on Biometrics on the Internet, 2004*, 2004, pp. 75–80.
- [165] N. Alyüz, B. Gökberk, L. Spreeuwers, R. Veldhuis, and L. Akarun, „Robust 3D face recognition in the presence of realistic occlusions“, in *2012 5th IAPR International Conference on Biometrics (ICB)*, IEEE, 2012, pp. 111–118.

- [166] P. J. Besl and N. D. McKay, „Method for registration of 3-D shapes“, in *Sensor fusion IV: control paradigms and data structures*, Spie, vol. 1611, 1992, pp. 586–606.
- [167] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn, „A region ensemble for 3-D face recognition“, *IEEE Transactions on Information Forensics and Security*, 3, no., pp. 62–73, 2008.
- [168] H. T. Tanaka, M. Ikeda, and H. Chiaki, „Curvature-based face surface recognition using spherical correlation. principal directions for curved object recognition“, in *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, IEEE, 1998, pp. 372–377.
- [169] T. Kurita and P. Boulanger, „Computation of Surface Curvature from Range Images Using Geometrically Intrinsic Weights.“, in *MVA*, 1992, pp. 389–392.
- [170] B. Gokberk, H. Dutagaci, A. Ulas, L. Akarun, and B. Sankur, „Representation plurality and fusion for 3-D face recognition“, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 38, no., pp. 155–173, 2008.
- [171] H. J. Seo and P. Milanfar, „Face verification using the lark representation“, *IEEE Transactions on Information Forensics and Security*, 6, no., pp. 1275–1286, 2011.
- [172] A. Rakotomamonjy, F. Bach, S. Canu, *et al.*, „SimpleMKL“, *Journal of Machine Learning Research*, 9, no., pp. 2491–2521, 2008.
- [173] G. R. Lanckriet, T. De Bie, N. Cristianini, M. I. Jordan, and W. S. Noble, „A statistical framework for genomic data fusion“, *Bioinformatics*, 20, no., pp. 2626–2635, 2004.
- [174] G. R. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. I. Jordan, „Learning the kernel matrix with semidefinite programming“, *Journal of Machine learning research*, 5, no., pp. 27–72, 2004.
- [175] B. K. Sriperumbudur, K. Fukumizu, and G. R. Lanckriet, „Universality, Characteristic Kernels and RKHS Embedding of Measures.“, *Journal of Machine Learning Research*, 12, no., 2011.
- [176] *MATLAB version 9.10.0.1613233 (R2021a)*, The Mathworks, Inc., Natick, Massachusetts, 2021.
- [177] G. Bradski, „The OpenCV Library“, *Dr. Dobb's Journal of Software Tools*, no., 2000.
- [178] C.-C. Chang and C.-J. Lin, „LIBSVM: A library for support vector machines“, *ACM Transactions on Intelligent Systems and Technology*, 2, no., 27:1–27:27, 3 2011, Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [179] R.-E. Fan, P.-H. Chen, C.-J. Lin, and T. Joachims, „Working set selection using second order information for training support vector machines.“, *Journal of machine learning research*, 6, no., 2005.
- [180] M. S. Pepe, „Receiver operating characteristic methodology“, *Journal of the American Statistical Association*, 95, no., pp. 308–311, 2000.

-
- [181] F. Bashar, A. Khan, F. Ahmed, and M. H. Kabir, „Robust facial expression recognition based on median ternary pattern (MTP)“, in *Electrical Information and Communication Technology (EICT), 2013 International Conference on*, IEEE, 2014, pp. 1–5.
- [182] P. Nair and A. Cavallaro, „SHREC’08 entry: Registration and retrieval of 3D faces using a point distribution model“, in *2008 IEEE International Conference on Shape Modeling and Applications*, IEEE, 2008, pp. 257–258.